

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:48:07 ; Search time 7965.97 Seconds

(without alignments)
10902.766 Million cell updates/sec

Title: US-10-024-370-1

Perfect score: 2123
Sequence: 1 ctccgagcgsgagctcggtgatc.....gacctcatgcctcggaattc 2123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext:1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_ph: *
7: gb_pl: *
8: gb_pr: *
9: gb_ro: *
10: gb_ry: *
11: gb_ry: *
12: gb_ry: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_fun: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_ph: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_ry: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_hum: *
32: em_hcg_inv: *
33: em_hcg_mus: *
34: em_hcg_mus: *
35: em_hcg_pln: *
36: em_hcg_rtd: *
37: em_hcg_mam: *
38: em_hcg_vrt: *
39: em_hcg_hum: *
40: em_hcg_mus: *
41: em_hcg_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2123	100.0	2123	1	CG117592	Y17592 Corynebacte
2	2123	100.0	2123	6	AR202319	AR202319 Sequence
3	2123	100.0	2123	6	BD004741	BD004741 Replicabl
4	2111	99.4	332050	1	AP005276	AP005276 Corynebac
5	2111	99.4	349980	6	AX127145	AX127145 Sequence
6	1575	74.2	1575	6	AX066443	AX066443 Sequence
7	1473	69.4	1473	6	AR202320	AR202320 Sequence
8	1473	69.4	1473	6	AX121013	AX121013 Sequence
9	1473	69.4	1473	6	BD004742	BD004742 Replicabl
10	1473	69.4	1473	6	BD163130	BD163130 Novel pol
11	691.4	32.6	300750	1	AP005217	AP005217 Corynebac
12	515	24.3	3013	1	CG117592	CG117592 Corynebacte
13	312.6	14.7	17910	1	AE006979	AE006979 Mycobacte
14	312.6	14.7	37630	1	MTCY31	MTCY31 Mycobacteri
15	312.6	14.7	327650	1	BX248337	BX248337 Mycobacteri
16	164.4	7.7	342300	1	MLEPRTN8	MLEPRTN8 Mycobacte
17	158.4	7.5	10029	1	AE006634	AE006634 Brucella
18	156.8	7.4	10029	1	AE014498	AE014498 Brucella
19	148.6	7.0	260050	1	SME591782	AL591782 Sinorhizo
20	142	6.7	36063	1	MLU15184	U15184 Mycobacteri
21	140.4	6.6	332635	1	AP003005	AP003005 Mesorhizo
22	136.6	6.4	298900	1	AP005937	AP005937 Bradyrhiz
23	132.6	6.2	294250	1	AP001517	AP001517 Bacillus
24	131	6.2	354	6	AX123526	AX123526 Sequence
25	131	6.2	354	6	BD165643	BD165643 Novel pol
26	128.4	6.0	7697	1	AE007944	AE007944 Agrobacte
27	128.4	6.0	9508	1	AE008976	AE008976 Agrobacte
28	124.8	5.9	6908	1	SPU59237	US9237 Synecococc
29	122.8	5.8	4776	1	AB031231	AB031231 Pseudomon
30	122.2	5.8	11976	1	AE004735	AE004735 Pseudomon
31	122.2	5.8	14401	1	AE001518	AE001518 Helicobac
32	120.4	5.7	304454	1	AE016956	AE016956 Enterococ
33	120.4	5.7	10740	1	AE012365	AE012365 Xanthomon
34	120.2	5.7	311321	1	NMA322491	AL162754 Neisseria
35	119.2	5.6	1053	6	BD092453	BD092453 Identific
36	119.2	5.6	14160	1	AE000604	AE000604 Helicobac
37	119	5.6	3777	1	AF139661	AF139661 Azospiril
38	118.8	5.6	300975	1	AE017013	AE017013 Bacillus
39	118.8	5.6	301200	1	AP005374	AP005374 Thermosyn
40	118.6	5.6	10377	1	AE002422	AE002422 Neisseria
41	118.6	5.6	349980	6	AX044030	AX044030 Sequence
42	117.2	5.5	291804	1	AE017039	AE017039 Bacillus
43	116.6	5.5	13369	1	AE006013	AE006013 Caulobact
44	113.4	5.3	300363	1	AE016781	AE016781 Pseudomon
45	110.8	5.2	14704	1	AE001970	AE001970 Deinococc

ALIGNMENTS

RESULT 1	CG117592							
LOCUS	CG117592	2123 bp	DNA	linear	BCT 26-JUN-2001			
DEFINITION	Corynebacterium glutamicum accDA gene.							
ACCESSION	Y17592							
VERSION	Y17592.1	GI:14572581						
KEYWORDS	accDA gene; acetyl-CoA carboxylase; carboxyltransferase.							
SOURCE	Corynebacterium glutamicum							
ORGANISM	Corynebacterium glutamicum							
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;							
	Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.							
REFERENCE	1 Kim, K., Peters-Wendisch, P.G., Salm, H. and Eikmann, B.J.							
AUTHORS	unpublished							
JOURNAL	2 (bases 1 to 2123)							
REFERENCE								

AUTHORS Peters-Wendisch, P.G.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1998) P.G. Peters-Wendisch, University of California at Berkeley, Dept. Plant and Microbial Biology, Kuecu Lab, 111 Koshland Hall, Berkeley CA 94720, USA

COMMENT Related sequence x66112.

FEATURES Location/Qualifiers

1..2123

/organism="Corynebacterium glutamicum"

/mol_type="genomic DNA"

/strain="ATCC 13032"

/db_xref="taxon:1718"

508..1983

/gene="naccDA"

508..1983

/gene="naccDA"

/codon_start=1

/transl_table=1

/product="putative carboxyltransferase subunit of

acetyl-CoA carboxylase"

/protein_id="CAC42827.1"

/db_xref="GI:14572582"

/translation="MEKRFPTMVGMEHTSLALTLIDSLDPSISWNETPYQYDNLQ

GYAETLBRARSKACDSVITGEETVEGI PAVILSDPSFLGSLGTVASVIRKAIH

RATLKLPLIVSPASGGARMOEDNRA FVMWVSITAAVORHEALPLIVLRNTMGS

AMLSWSSGHLTPAPRQAOTGFLGPRVLELTGHALPDQQAENLVKTVGIDITVSP

LQLRRAVAKTIKVIQPEVATDRFSPFTTGVVALPMEALARSDFQRIGIMETLGA

DVNLGSRAGALSPAVRVALARIGRVRVILIGDRRTLSPOELRRRISLAREL

NLPYVSIIDTSAGELSOAEELGIASSIARTLSKIDAPLPVSVIIGGGGGALAM

LPLADLVAAENAMLSALPPEGASAILFEDTTHAAEILIERQCVQAHALLISQGLIDGIVA

ETHEVEEILGTISNALSLELDNPERARDRSRFRFERLAQ"

BASE COUNT 460 a 574 c 592 g 497 t

ORIGIN

Query Match 100.0%; Score 2123; DB 1; Length 2123;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGAGCGGGAGTGGGTATGGCCACTCTTAAGCAATGCCGGTTTAAATTAAGCA 60
Db 1 CTCGAGCGGGAGTGGGTATGGCCACTCTTAAGCAATGCCGGTTTAAATTAAGCA 60
QY 61 CTTATATGTTCTCCACCATCTGCGCCAGACGAGATGTCGATTCACAGCTA 120
Db 61 CTTATATGTTCTCCACCATCTGCGCCAGACGAGATGTCGATTCACAGCTA 120
QY 121 AACGTGAATGTGAATCTTAACCTCAATGTCATGCGATGCGATTTGAAACCTC 180
Db 121 AACGTGAATGTGAATCTTAACCTCAATGTCATGCGATGCGATTTGAAACCTC 180
QY 181 ACTGCCCCCAATATTCTTAACCTTAAGTAAGTGTCTTAACTGATTTATTAAGT 240
Db 181 ACTGCCCCCAATATTCTTAACCTTAAGTAAGTGTCTTAACTGATTTATTAAGT 240
QY 241 TCCGATCTACCCCTCTTAAACCCGGAATACCCCTTTTGCAAGATGCAACACACA 300
Db 241 TCCGATCTACCCCTCTTAAACCCGGAATACCCCTTTTGCAAGATGCAACACACA 300
QY 241 TCCGATCTACCCCTCTTAAACCCGGAATACCCCTTTTGCAAGATGCAACACACA 300
Db 241 TCCGATCTACCCCTCTTAAACCCGGAATACCCCTTTTGCAAGATGCAACACACA 300
QY 301 GTGCAATGATTAACGGGGCTTCAACGTCACATTCGTCCGGTTTAAAGCTATGTTCCGG 360
Db 301 GTGCAATGATTAACGGGGCTTCAACGTCACATTCGTCCGGTTTAAAGCTATGTTCCGG 360
QY 361 ACGTCTAGGCAAAAAGTAGTTTGTGAGATGAAAGCAATACCGTCAATTTTAAAGCA 420
Db 361 ACGTCTAGGCAAAAAGTAGTTTGTGAGATGAAAGCAATACCGTCAATTTTAAAGCA 420
QY 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGTATGAGAGACATTGGA 480
Db 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGTATGAGAGACATTGGA 480
QY 481 ATTAATTAACAAAGCCATTTTTCGGCCGTGAGAGAGCGTTTCCGACTATGGTGTGGGC 540
Db 481 ATTAATTAACAAAGCCATTTTTCGGCCGTGAGAGAGCGTTTCCGACTATGGTGTGGGC 540

QY 541 ATGAACACACTTCAGATTGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCAT 600
Db 541 ATGAACACACTTCAGATTGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCAT 600
QY 601 TCTTGAATGAACCTCCCAATATGACAACTCAATCAAGGCTATGACAGACTTTGAG 660
Db 601 TCTTGAATGAACCTCCCAATATGACAACTCAATCAAGGCTATGACAGACTTTGAG 660
QY 661 CGGGCTCGAAGCAAGGCAATATGCGATGATCGGTAAATTAATCTGAGAAAGGACCGTGAG 720
Db 661 CGGGCTCGAAGCAAGGCAATATGCGATGATCGGTAAATTAATCTGAGAAAGGACCGTGAG 720
QY 721 GGCATTCGGTAGCCGTTATTTTGTCCGATTTTCTCTCCGCGGTTCTTTGGGACG 780
Db 721 GGCATTCGGTAGCCGTTATTTTGTCCGATTTTCTCTCCGCGGTTCTTTGGGACG 780
QY 781 GTCGCGTCGGTGGCATCATGAAGGCGATTACCGGCGCACAGAGCTGAATCTCCACTG 840
Db 781 GTCGCGTCGGTGGCATCATGAAGGCGATTACCGGCGCACAGAGCTGAATCTCCACTG 840
QY 841 CTGGTCTCCCTGCTTCCGGTGGTGGCGCATGCAAGAAACAATGAGCTTTTGTATG 900
Db 841 CTGGTCTCCCTGCTTCCGGTGGTGGCGCATGCAAGAAACAATGAGCTTTTGTATG 900
QY 901 ATGGTGTCCATTAACCGGCGCTGTCAGCGCTCACCGGAGGCGCATTTGCCGTTCTG 960
Db 901 ATGGTGTCCATTAACCGGCGCTGTCAGCGCTCACCGGAGGCGCATTTGCCGTTCTG 960
QY 961 TATTTGGCAATCCACGATGGGTGGCGCATGAGCTGTGGGATTCATGCGGATCTC 1020
Db 961 TATTTGGCAATCCACGATGGGTGGCGCATGAGCTGTGGGATTCATGCGGATCTC 1020
QY 1021 ACTTTTGGCAATCCCGGCGCGGAGATGATTTCTGGGTCTCTGGGTCTGCGGTGAGTTAAC 1080
Db 1021 ACTTTTGGCAATCCCGGCGCGGAGATGATTTCTGGGTCTCTGGGTCTGCGGTGAGTTAAC 1080
QY 1081 ACTGGGATGCGCTTCCAGACGAGTGTGACAGAGGCGGAAATTTGGTGAATCTGTGTG 1140
Db 1081 ACTGGGATGCGCTTCCAGACGAGTGTGACAGAGGCGGAAATTTGGTGAATCTGTGTG 1140
QY 1141 ATTGATGAATTTGTGTGCGCACTCAATTTGCGTGCAGCGGTGGCAAAACCTCAAGTT 1200
Db 1141 ATTGATGAATTTGTGTGCGCACTCAATTTGCGTGCAGCGGTGGCAAAACCTCAAGTT 1200
QY 1201 ATTGATGCGGTGAAGGCAAGGATGCTTTTCTCAACACTCTGCGGTGGCACTTCCG 1260
Db 1201 ATTGATGCGGTGAAGGCAAGGATGCTTTTCTCAACACTCTGCGGTGGCACTTCCG 1260
QY 1261 GTGATGAGGAGGATTTGGCGCTTCGTCGACCCGACAGAGGCTGGAATCGGGAGATTATG 1320
Db 1261 GTGATGAGGAGGATTTGGCGCTTCGTCGACCCGACAGAGGCTGGAATCGGGAGATTATG 1320
QY 1321 GAAACGTTGGGGGAGAGCTGTCGTCAGCTTCTGCTGCGGTGGCGCATTTAGGCCG 1380
Db 1321 GAAACGTTGGGGGAGAGCTGTCGTCAGCTTCTGCTGCGGTGGCGCATTTAGGCCG 1380
QY 1381 GCTGTGCGCTTGGCGCTGCGGCGCATGCGGGGCGGCGGTGTGATTTGGGCAAGAT 1440
Db 1381 GCTGTGCGCTTGGCGCTGCGGCGCATGCGGGGCGGCGGTGTGATTTGGGCAAGAT 1440
QY 1441 CGCGGCTTCAACGCTTGGGCGGAGAGCTGCTTGGCGGTGGCGCATTTGCGTGGCG 1500
Db 1441 CGCGGCTTCAACGCTTGGGCGGAGAGCTGCTTGGCGGTGGCGCATTTGCGTGGCG 1500
QY 1501 CGGAGCTTAAACCTGCGGATGCTGTCATCATGACACCTCCGCGCGGCAATTTGTCAG 1560
Db 1501 CGGAGCTTAAACCTGCGGATGCTGTCATCATGACACCTCCGCGCGGCAATTTGTCAG 1560
QY 1561 GCGGCTGAGAGGCTCGGATGCTGCAATGCTGCAATTTGCGGACCTTTGTCAGCTTATGAC 1620
Db 1561 GCGGCTGAGAGGCTCGGATGCTGCAATGCTGCAATTTGCGGACCTTTGTCAGCTTATGAC 1620

QY 1621 GCTCCCTCCCAACCGTTTCGATTAATGATCAGGCGCTGGCGGCTGGCC 1680
DB 1621 GCTCCCTCCCAACCGTTTCGATTAATGATCAGGCGCTGGCGGCTGGCC 1680
QY 1681 ATGCTGCCCCCGATCTGCTACGCGGCGGAAAGCGGTGCTGCCATTCGACCA 1740
DB 1681 ATGCTGCCCCCGATCTGCTACGCGGCGGAAAGCGGTGCTGCCATTCGACCA 1740
QY 1741 GAGGCGCTCGGCACTCTCTCCGCGACACCAACGCGGGAATCATAGAGCG 1800
DB 1741 GAGGCGCTCGGCACTCTCTCCGCGACACCAACGCGGGAATCATAGAGCG 1800
QY 1801 CAAGCGTGCAGCGCAGCAGCTTTTAAGCAGGCTTATCGACGCGATTCGCGCA 1860
DB 1801 CAAGCGTGCAGCGCAGCAGCTTTTAAGCAGGCTTATCGACGCGATTCGCGCA 1860
QY 1861 ACCGACATCTTTGTTGAAGAAATTCGCGCAGATGACAGCGCCTCTCCGAATGGAT 1920
DB 1861 ACCGACATCTTTGTTGAAGAAATTCGCGCAGATGACAGCGCCTCTCCGAATGGAT 1920
QY 1921 AACCAATCGGAGAGGCGGAGCGCAGTCCGCTTACAGATTTGAGGCTTAAAGCGAG 1980
DB 1921 AACCAATCGGAGAGGCGGAGCGCAGTCCGCTTACAGATTTGAGGCTTAAAGCGAG 1980
QY 1981 TAAAGAAATTAATGCGCTGATCAATGATGATGAACACAGGCTACGCGCAGACAGTGG 2040
DB 1981 TAAAGAAATTAATGCGCTGATCAATGATGATGAACACAGGCTACGCGCAGACAGTGG 2040
QY 2041 GTGGCGGAACTCTCAAGGCGGCTTGAAGCAGCTTCGCGGAATGCTGAGCTGACGAGTCC 2100
DB 2041 GTGGCGGAACTCTCAAGGCGGCTTGAAGCAGCTTCGCGGAATGCTGAGCTGACGAGTCC 2100
QY 2101 GCCGACCTTCATGCTCGGAATTC 2123
DB 2101 GCCGACCTTCATGCTCGGAATTC 2123

RESULT 2
LOCUS AR202319 2123 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6361986.
ACCESSION AR202319
VERSION AR202319.1 GI:20256858
KEYWORDS
SOURCE
ORGANISM
Unkown.
REFERENCE
1 (bases 1 to 2123)
111g,Y., Bikhanna,B., Eggeling,L., Sahm,H. and Mockel,B.
TITLE
Process for the preparation of L-amino acids by fermentation and
nucleotide sequences coding for the accDA gene
JOURNAL
Patent: US 6361986-A 1 26-MAR-2002;
FEATURES
Location/Qualifiers
1..2123
/organism="unkown"
BASE COUNT 460 a 574 c 592 g 497 t
ORIGIN

Query Match 100.0%; Score 2123; DB 6; Length 2123;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGAGCGGAGTGGTGTATGCGGCACTCTTAAGCATGCGGCTTTAAATAAGCA 60
DB 1 CTCGAGCGGAGTGGTGTATGCGGCACTCTTAAGCATGCGGCTTTAAATAAGCA 60
QY 61 CTTATATGTTTCTACACATCTGGCGAGCAGCAGAGTATGTTGTCGATCAAGCTA 120
DB 61 CTTATATGTTTCTACACATCTGGCGAGCAGCAGAGTATGTTGTCGATCAAGCTA 120
QY 121 AACGTGTAATGTGAAGTTAAGTAACTGACATGCAATGCAATGCGATTTGGAAAACTC 180
DB 121 AACGTGTAATGTGAAGTTAAGTAACTGACATGCAATGCAATGCGATTTGGAAAACTC 180

QY 181 ACTCCCCCAATATCTTAATCTTAACTTAAAGTAGTGTCTTACCTGCATTTATAAAGT 240
DB 181 ACTCCCCCAATATCTTAATCTTAACTTAAAGTAGTGTCTTACCTGCATTTATAAAGT 240
QY 241 TCCGATCTAACCCCTCTTTTACCCGAAATACCCCTTTTGCAAAAGTGCACAAACA 300
DB 241 TCCGATCTAACCCCTCTTTTACCCGAAATACCCCTTTTGCAAAAGTGCACAAACA 300
QY 301 GTGCAATAGTTAACGGGCTTTCACAGTCCATCTGTCCGCTTTAGGCTATGTTCCGG 360
DB 301 GTGCAATAGTTAACGGGCTTTCACAGTCCATCTGTCCGCTTTAGGCTATGTTCCGG 360
QY 361 AGGTGAGCAAAAGTATGTTTGTGAATGAAAGCAATATCCGATTTTAAAGCA 420
DB 361 AGGTGAGCAAAAGTATGTTTGTGAATGAAAGCAATATCCGATTTTAAAGCA 420
QY 421 TCGATAGCTTAATTTGAGGCTTAGATCTTCGCGCTCTAAATAGATGACAGACATTGCA 480
DB 421 TCGATAGCTTAATTTGAGGCTTAGATCTTCGCGCTCTAAATAGATGACAGACATTGCA 480
QY 481 ATTAAATTAAAGCAAGCATTTTTCGCGCGTGAAGACGTTTTCGATATGCTGCGGC 540
DB 481 ATTAAATTAAAGCAAGCATTTTTCGCGCGTGAAGACGTTTTCGATATGCTGCGGC 540
QY 541 ATGGAACACATTCAGCATTTGACGCTCATGACTGGTTTGGACCTGACACTTCAAT 600
DB 541 ATGGAACACATTCAGCATTTGACGCTCATGACTGGTTTGGACCTGACACTTCAAT 600
QY 601 TCTTGAATGAAGACTCCCAATATGACAACTCATCAAGGCTATGACAGACCTTGGAG 660
DB 601 TCTTGAATGAAGACTCCCAATATGACAACTCATCAAGGCTATGACAGACCTTGGAG 660
QY 661 CGGAGCTGAAAGCAAGGCAAAATGCGATGATGCTAATTAATCTGGAAGGCAAGCTGGAG 720
DB 661 CGGAGCTGAAAGCAAGGCAAAATGCGATGATGCTAATTAATCTGGAAGGCAAGCTGGAG 720
QY 721 GGCATTCGCGTAGGCGGTATTTTGTCCGATTTTCTTCTGCGGCTTCTTGGGACG 780
DB 721 GGCATTCGCGTAGGCGGTATTTTGTCCGATTTTCTTCTGCGGCTTCTTGGGACG 780
QY 781 GTGCGCTGAGTGCAGTATGAAAGGCAATTCACGCGGCAAGACGTGAATCTCCACAG 840
DB 781 GTGCGCTGAGTGCAGTATGAAAGGCAATTCACGCGGCAAGACGTGAATCTCCACAG 840
QY 841 CTGATCTCCCTGCTTCCGATGTGTCGCGCATGACAGAAACAATGAGCTTTTGTCAAG 900
DB 841 CTGATCTCCCTGCTTCCGATGTGTCGCGCATGACAGAAACAATGAGCTTTTGTCAAG 900
QY 901 ATGCTGTCAATPACCGGCGCTGTGCAAGCTGACCGGAGGCGCATTTCCGCTGTGG 960
DB 901 ATGCTGTCAATPACCGGCGCTGTGCAAGCTGACCGGAGGCGCATTTCCGCTGTGG 960
QY 961 TATTGGCAATCCACAGATGGTGGCGCAATGCGCTGTGGGCTTCAATCGGAGATCTC 1020
DB 961 TATTGGCAATCCACAGATGGTGGCGCAATGCGCTGTGGGCTTCAATCGGAGATCTC 1020
QY 1021 ACTTTTGGGAACCGGCGGCAAGTAGTGTCTCGGCTGCTGCGGTGGAAGTTAAC 1080
DB 1021 ACTTTTGGGAACCGGCGGCAAGTAGTGTCTCGGCTGCTGCGGTGGAAGTTAAC 1080
QY 1081 ACTGGGCAATGCGCTTCCAGAGCGTGTGCAAGCGGCGGAATTTGTGAAAACTGTGTG 1140
DB 1081 ACTGGGCAATGCGCTTCCAGAGCGTGTGCAAGCGGCGGAATTTGTGAAAACTGTGTG 1140
QY 1141 ATTGATGGAATTGTGTGCGCACTCCAAATGGTGCAGCGGTGGAAGAAACCTCAAGTT 1200
DB 1141 ATTGATGGAATTGTGTGCGCACTCCAAATGGTGCAGCGGTGGAAGAAACCTCAAGTT 1200
QY 1201 ATTGAGCGGATGAGGCAAGCATGCTTTTCTCAACAACTCTCGGCTGGGCACTTCCG 1260
DB 1201 ATTGAGCGGATGAGGCAAGCATGCTTTTCTCAACAACTCTCGGCTGGGCACTTCCG 1260

QY	126	GTANTGAGAGCGATTGCGCGCTTTCGTGATACCCGCGAGAGCGCTCGGATCGGGAGGATTATG	1320
Db	1261	GTGATGAGAGCGGATTGGCGCTTCTCGTATCCCGCAGAGCGCTCGAATCGGGGAATTATG	1320
QY	1321	GAAGCGTTGGGGGAGAGCGTGTGTCAAGCTTTCTGTGTGCGCTGTCTGCGCATTTAGACCCG	1380
Db	1321	GAAGCGTTGGGGGAGAGCGTGTGTCAAGCTTTCTGTGTGCGCTGTCTGCGCATTTAGACCCG	1380
QY	1381	GCTGTGCGCGTTGCGCTGTGCGCGCATCGGGGGGCGGGCCCGTGGTGTGCTGATTTGGGCGAGGAT	1440
Db	1381	GCTGTGCGCGTTGCGCTGTGCGCGCATCGGGGGGCGGGCCCGTGGTGTGCTGATTTGGGCGAGGAT	1440
QY	1441	CGCGCGTTCAAGCTTTGGGCGCGAGAGCGTGTGCTTTGGCGCGTGTGCGATTTGCTGTGCG	1500
Db	1441	CGCGCGTTCAAGCTTTGGGCGCGAGAGCGTGTGCTTTGGCGCGTGTGCGCATTTGCTGTGCG	1500
QY	1501	CGCGAGCTTAACCTTGCGGATGTGTCTCATTCAGACACTCTCGGCGCCGGAATTTGTCTGAG	1560
Db	1501	CGCGAGCTTAACCTTGCGGATGTGTCTCATTCAGACACTCTCGGCGCCGGAATTTGTCTGAG	1560
QY	1561	CGCGCTGAGAGAGCTCGCGCATCGCAAGCTTCGATTTGGCGCGCACTTGTCCAGACTTATCGAC	1620
Db	1561	CGCGCTGAGAGAGCTCGCGCATCGCAAGCTTCGATTTGGCGCGCACTTGTCCAGACTTATCGAC	1620
QY	1621	GCTTCCCTTCCCAACCGTTTCGATTAATTGTCAAGGCGTTTGGCGGTGGCGCGCTGTGACC	1680
Db	1621	GCTTCCCTTCCCAACCGTTTCGATTAATTGTGTCAAGGCGTTTGGCGGTGGCGCGCTGTGACC	1680
QY	1681	ATGTCGACCCGCGGATCTGTGTCTAGCGCGGCGGAAGAGGTGGCGTCCGCAATTGCGCACCA	1740
Db	1681	ATGTCGACCCGCGGATCTGTGTCTAGCGCGGCGGAAGAGGTGGCGTCCGCAATTGCGCACCA	1740
QY	1741	GAGGCGCGCTCGGCCATCTCTTCCGCGACACCAACGCGCGCGGAAATCATTAGACGGA	1800
Db	1741	GAGGCGCGCTCGGCCATCTCTTCCGCGACACCAACGCGCGCGGAAATCATTAGACGGA	1800
QY	1801	CAAGCGGTGACGCGGCGGACGCACTTTTAAAGCAAGGCGTTATCGAGGGATGCTCGCCGAA	1860
Db	1801	CAAGCGGTGACGCGGCGGACGCACTTTTAAAGCAAGGCGTTATCGAGGGATGCTCGCCGAA	1860
QY	1861	ACCGGACCTTGTGTGAAGAAATTTCTCGGCAATCAGCAACGCGCTTCTCGAATTGGAT	1920
Db	1861	ACCGGACCTTGTGTGAAGAAATTTCTCGGCAATCAGCAACGCGCTTCTCGAATTGGAT	1920
QY	1921	AACCAATCCGAGAGGGCGGAGCGACGCACTGTCTTCAACGATTTTGAACGTTTAGCGGAG	1980
Db	1921	AACCAATCCGAGAGGGCGGAGCGACGCACTGTCTTCAACGATTTTGAACGTTTAAACGAG	1980
QY	1981	TAAAGAAATATTATGCGTGTATCAATTCGATGTATGAACACAGGGTACGCGCGAGACAGTGG	2040
Db	1981	TAAAGAAATATTATGCGTGTATCAATTCGATGTATGAACACAGGGTACGCGCGAGACAGTGG	2040
QY	2041	GTGCGCGGAACTTCAGGGCCGTTAAGCAGCGCTGTGCGGAATGTCACTGACGACGTC	2100
Db	2041	GTGCGCGGAACTTCAGGGCCGTTAAGCAGCGCTGTGCGGAATGTCACTGACGACGTC	2100
QY	2101	GCCGACCTTCATGCTCGGAATTC	2123
Db	2101	GCCGACCTTCATGCTCGGAATTC	2123

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 2123)	Tilke,I., Eggerting,L., Eickmans,B., Zamu,H. and Meck,V.		
		Replitec DNA, amino acid sequence, Corynebacterium microorganism,		
		shuttle vector, and process for producing L-amino acid		
	Patent: JP 2001008693-A-1	16-JAN-2001,		
	DRUGUSA HUELS AG,FORSCHUNGSZENTRUM JOEBLICH GMBH			
	OS	Corynebacterium glutamicum		
	PN	JP 2001008693-A/1		
	PD	16-JAN-2001		
	PF	24-MAY-2000 JP 2000153547		
	PR	27-MAY-1999 DE 19924365.4		
	PI	IVENNE TILKE,LOTHAR EGGERING,BERNHARD EICKMANS,HERMANN ZAMU,		
	PI	VERLYNN MECKER		
	PC	C12N15/09,C12N1/21,C12P13/04,C12P13/06,C12P13/06,C12P13/08, PC		
		C12P13/08,		
	PC	C12P13/12,C12P13/20,C12P21/02/(C12N15/09,C12R1:15),(C12N1/21,		
	PC	C12R1:15),		
	PC	(C12P13/04,C12R1:15),(C12P13/06,C12R1:15),(C12P13/08,C12R1:15), PC		
		(C12P13/12,C12R1:15),(C12P13/20,C12R1:15),(C12P13/02,C12R1:15), PC		
		C12N15/00,		
	PC	(C12N15/00,C12R1:15)		
	CC			
	FH	Key	Location/Qualifiers	
	FT	gene	(508). .(1980).	
		location/Qualifiers		
		1..2123		
		/organism="Corynebacterium glutamicum"		
		/mol_type="genomic DNA"		
		/db_ref="taxon:1718"		
BASE COUNT		460 a	574 c	592 g 497 t
ORIGIN				

Query Match	100.0%;	Score 2123;	DB 6;	Length 2123;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2123;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CTCGAGCGGAGGTGCGTGAATCGGCCACTCTCTAAGCAATGCGGGCTTTAAATTAAGCAA	60	
Db	1	CTGAGGCGGGAGTGGTGATGCGGCACCTCTCTAAGCAATGCGGGCTTTAAATTAAGCAA	60	
Qy	61	CTTATATGTTTCTTCACCAATCTGGCCGACGACCAAGATATGTTGCGATCACAGCTA	120	
Db	61	CTTATATGTTTCTTCACCAATCTGGCCGACGACCAAGATATGTTGCGATCACAGCTA	120	
Qy	121	AACGTGTGAATGTGAATTAACCTTAACCTCACTGCAATGGCATGGATGGCAATTTGGAAATC	180	
Db	121	AACGTGTGAATGTGAATTAACCTTAACCTCACTGCAATGGCATGGCAATTTGGAAATC	180	
Qy	181	ACTGCCCCCAATATCTTAACCTTAACTTAAAGTAGTGTTTACCTGCATTTATTAAGA	240	
Db	181	ACTGCCCCCAATATCTTAACCTTAACTTAAAGTAGTGTTTACCTGCATTTATTAAGA	240	
Qy	241	TCCGATCTACCCGCTCTTTAACCCCGAAATACCCTTTGCAGAAATGTGCAACACACA	300	
Db	241	TCCGATCTACCCGCTCTTTAACCCCGAAATACCCTTTGCAGAAATGTGCAACACACA	300	
Qy	301	GTGCAATATGTTAACGGGCTTCAACGTCACATTTCTGCGGTTTAGGCTATGTTGGG	360	
Db	301	GTGCAATATGTTAACGGGCTTCAACGTCACATTTCTGCGGTTTAGGCTATGTTGGG	360	
Qy	361	ACGCTTAGGCAAAAAGTAGTTTGTGAGATGAAGAAACGATATCCGTCATTTTAAACCAA	420	
Db	361	ACGCTTAGGCAAAAAGTAGTTTGTGAGATGAAGAAACGATATCCGTCATTTTAAACCAA	420	
Qy	421	TGCGATACCTTAATTTGGGCTTAGATCTTTCGCGCTCTAAATATAGTATGACAGACATTGCA	480	
Db	421	TGCGATACCTTAATTTGGGCTTAGATCTTTCGCGCTCTAAATATAGTATGACAGACATTGCA	480	
Qy	481	ATTATATTAACAAAGCCATTTTTCGGCGGTGAGAGAGGTTTTCGACTATGCTGTGGGCG	540	
Db	481	ATTATATTAACAAAGCCATTTTTCGGCGGTGAGAGAGGTTTTCGACTATGCTGTGGGCG	540	


```

/codon_start=1
/transl_table=11
/product="Carbon starvation protein, predicted membrane
protein"
/protein_id="BAB98047.1"
/db_xref="GI:21323420"
/translation="MKKSPKMSPTARVQLVFGVIAVGWGAIAFSRGETINSVWLVL
AAGSVLIAFSFYARLIEKVKKPKDQATPAEIVNDGKDVPTDRRLFGHHPAALA
GAGPLVGPMVAOMGYLPGTLITLIGVIFAGVODVYLWVSTRGRGRSGOMVRDM
GTGAGAGIATISIMITITLVALIVNALADSPKGVFSITMTITPILAPMGVYRLRL
RPGVETSIIGVALLLAIIVAGMVDTSWGEWFTWSKTTLLALIGIMAILP
WMLLAPRDYLSITPMKIGVIGLLAVGLFAPREVQMSVTSFALGNGVPVSGSLFEP
LFTIACGASGFPALISSGTPPLVKEKESQRMIGYGMIMESFVAMALITVILID
RHLTFSMNAPLATGDPATAAEWNSIGLTGADITPEOLSEAESVGESTVIRTCG
APLAFGMSSETISGFIGAGKAKAWMHPALIMEALFITTYDAGTAVRAPMMDTTLG
VPGRLRRKDSWYTNMISTYFVVALMGAILIMKVTPLPGCINLFPULFGIANDLLA
IALVLIVVVVKGLYKRAMIPAPPLAMDLVITVTASMOKIFHSDPAIWAQVANR
DASQGLTERFGAASPEAIDVIRNTWIOGLSTLFVLVLVVGAAIIVCISKIRAR
AAGPLPTEEPDTESEFFAFTGFLASSRDXEYQAMWDERYPGAPVSSGGH"
2522..2725
/gene="Cg10655"
2522..2725
/gene="Cg10655"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98048.1"
/db_xref="GI:21323421"
/translation="WMALTHALMKIPRAVMWYLTLMGDRAYSRYVHLKHHHDAPL
PTREYVRAKADQDANPGARCC"
complement(3092..3703)
/gene="Cg10656"
complement(3092..3703)
/gene="Cg10656"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98049.1"
/db_xref="GI:21323422"
/translation="MSNVINILYMQAHSHTQVNLIPSSNATQIILATSSSLCCSQ
NRNVISWGVFRSEERVMEIKTLHRSVSGRSFPGITSMQTEKRCPTTIGMPPDC
HRTGRANWSTQHSCLHRCIVDDVDFHLSVRNLSLWVGCMASDLVSQVLAPELFR
RTVGAFVNVHGMVTPSORTVNAHTNSLSKSNVGHCKSCVIL"
3750..4706
/gene="Cg10657"
3750..4706
/gene="Cg10657"
/codon_start=1
/transl_table=11
/product="Unclassified protein involved in propionate
catabolism"
/protein_id="BAB98050.1"
/db_xref="GI:21323423"
/translation="MWNITLYQALHTTATROSKKAISSWKAFAFAPAGKMSIE
AYDAMRGEGAPSIWEGEDGVIMLISGDHITPLPAEGAKRALILTPTKESHA
EYQOAPIDILARSWGEKLAQGLRDVDSVLHSTHTHYIGTGSANDPOKPPDS
RETDHSIMYIFAVLEDRAMHERSVAPRAHREITELMNKISTVEDEMPRRYS
VDRPKAFGARAVITPKDGYVBELAVADNHPGAPPRAREQITQKFRILAEVASE
KEQDRFLDAQRHLEDELSELNTELDADILAKAPVPEGLF"
4715..5629
/gene="Cg10658"
4715..5629
/gene="Cg10658"
/codon_start=1
/transl_table=11
/product="PEP phosphonomutase and related enzymes"
/protein_id="BAB98051.1"
/db_xref="GI:21323424"
/translation="MFSSAVAPTRERKALAAALAPRIAMPGFSPALARATQEAF
EGYVSGAVVAADLALPDIGITLFEVANSRQJARTYDLPVLVADDTGFEPSAR
TVELEDAGVAGCHLEQVNVKRCGHLDGEKVGDTMVRRIAAVNERDEQCVIC
RTDAAGVEGIDSAIERAKAYADAGAMI FTEALYSPADFEKFAAVDIPILANTEG
KTELPAQLLEDIGYNAVITYPVTLLIAMGVQEQALDIANTGIDTMDVDRMQRSRL

```

```

gene
5658..6809
/gene="Cg10659"
5658..6809
/gene="Cg10659"
/EC_number="4.1.3.7"
/translation="PF00285:Citrate synthase"
/codon_start=1
/transl_table=11
/product="Citrate synthase"
/protein_id="BAB98052.1"
/db_xref="GI:21323425"
/translation="MSSATTTDVKGLYGIADYTAVSXVMPETNSITRYGAVEDLV
EANSFEFVFLHNGELPTAQOAEFRNERSYRSIDAGLISLHLPKPAHMDVR
TAVSGTQKSEVFTTDSERI RVGHITLALQPLMVLAMDRIIRRGDIIAPDSKSYA
ENLISVYRGTSPPASPNADVDPEKSLILYVHSGFNASTPARITYTSKSYGSI
TGAIGLKGPHGAGNPFVHNTMLAIDDPKAAWAMNNALDNKQVWGFGRYYKRD
SRVPSMEKSFRELARSHDEKQVAMNEMMDARTGAKPNLDFPAGPYHLLGPPV
DFFTPLFVIRVAGWTAHIVEQYENSLIRPLSEYNGEOREVAPIEKR"
complement(6846..7040)
/gene="Cg10660"
complement(6846..7040)
/gene="Cg10660"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98053.1"
/db_xref="GI:21323426"
/translation="MYFDVSQDVDELHPSFFELVADQALAHRCAPHYVWREAVA
KHVLPQSQKNWVTRMLHNG"
complement(7129..8172)
/gene="Cg10661"
complement(7129..8172)
/gene="Cg10661"
/EC_number="1.1.1.82"
/translation="PF02615:Malate/L-lactate dehydrogenase"
/codon_start=1
/transl_table=11
/product="Malate/L-lactate dehydrogenases"
/protein_id="BAB98054.1"
/db_xref="GI:21323427"
/translation="MPEVYNAQQLTVLCTDILTKTVGPAADAHVGDLSIVQADLMCH
RSHGVRLPVMYVRLSHSGAMTTHAHVEVNDLCAVALDGHNGIQVLDHARKAVT
PAMFGIAGVANSNHFPTAMYTRKAAOGCVSLITNNSPAMAPWGRKERIGIN
PMSIAPRGETAVDIDANVAVARGIYHARQNMNIPETWALTSGAFTDPDAVAF
GVYLPAGHKGYAISMMDVLSVLSQSHSTVHPRPYPTPGAGCHIPIADVAAP
RDPQDPDDALSDLVGEVKSPTKQANTBELTFPESEDRARHNSAHGISLPEKTMEL
QELAINHVYTR"
complement(8265..8987)
/gene="Cg10662"
complement(8265..8987)
/gene="Cg10662"
/translation="PF00392:Bacterial regulatory proteins, gntR family"
/codon_start=1
/transl_table=11
/product="Transcriptional regulators"
/protein_id="BAB98055.1"
/db_xref="GI:21323428"
/translation="WYQDAQASPAIRKADQVYQIRREIEDGTLNPGQMSVWLVE
HTGASRTPVADALRLADELILPEPQAPWSPLIRHDKLFEFRRIVEYALBEI
SVGASRSPRIFGESTLADELRELENSADDADTFDRRLTSKDFDLVANTNOPIG
KSILSKPHTRLRIITASHDPAHROSVOEHIMCSAVASGDRSAGACRQHLHVE
KSILPILINADSGSGGIDIRS"
complement(9094..9798)
/gene="Cg10663"
complement(9094..9798)
/gene="Cg10663"
/codon_start=1
/transl_table=11
/product="Hypothetical membrane protein"
/protein_id="BAB98056.1"
/db_xref="GI:21323429"
/translation="MRFSRVLPALLITTAVSIPASAAVLTADTDKELCIASNTDOSA

```

VTTFMNSIDSVREORLDELDAODPEIKAIIESYIAQDNAPFLAELOVRLDAISSGE
GLAMLLPDPPLTADPNASESFKTEYEDAKDIIISGSSDPSADIVLSQCAQATGTR
TIRAEVAFADRTDDYNEESQTLKEDFOCIDAIIDARPPIPOYLIIIGAIINAVIVL
GIRAWNSRKOSKHSO"
complement(9878) .12094)
/gene="Cg10664"
complement(9878) .12094)
/BC number="1.1.1.42"
/note="TRIGR00178:monomer_idh: isocitrate dehydrogenase,
NADP-dependent"
/codon_start=1
/transl_table=11
/product="Monomeric isocitrate dehydrogenase (NADP+)"
/protein_id="BAB98057.1"
/db_xref="GI:21323430"
/translation="MAKTIWTRTDEAPLTIATYSIKRVFAFAFATAGIEVEYETLISAG
RLIAGPERLETDQKVNALAEGLBELAPEANIIKLPIISASVPLKAIKELIDOG
YDIPLPDVAITDEKDIILARYNAVKGSVNVLEHNSDRAPIAVKNFVQIFPHRM

Query Match 99.4%; Score 2111; DB 1; Length 332050;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CTCGAGCGGAGCTCGGTGATCGGCCACTCTTAAGCAATGCCGCTTTAAATAAGCA 60
Db 213670 CTCGAGCGGAGCTCGGTGATCGGCCACTCTTAAGCAATGCCGCTTTAAATAAGCA 213611
Qy 61 CTTAATATGTTTCTACACACATCTGGCCGACGACGAGAAATGTTTGTGATCAGCTA 120
Db 213610 CTTAATATGTTTCTACACACATCTGGCCGACGACGAGAAATGTTTGTGATCAGCTA 213551
Qy 121 AACGTGTAATGTAAGTACCTAATCTCAATTGCAATGCGATGCGATTTGGAACATC 180
Db 213550 AACGTGTAATGTAAGTACCTAATCTCAATTGCAATGCGATGCGATTTGGAACATC 213491
Qy 181 ACTCCCCCAATATCTTAACCTTAATAAGTAGTGTGTTTAACTGCAATTTATAAGT 240
Db 213490 ACTCCCCCAATATCTTAACCTTAATAAGTAGTGTGTTTAACTGCAATTTATAAGT 213431
Qy 241 TCCCGATCTACCCCTCTTTACCCCGAAATACCCCTTTTGCAAGATGCAACACACA 300
Db 213430 TCCCGATCTACCCCTCTTTACCCCGAAATACCCCTTTTGCAAGATGCAACACACA 213371
Qy 301 GTGCAATAGTTAAGCGGCTTCAAGTCAACATCTGTCGCGTTTTAGCTATGTTCCGG 360
Db 213370 GTGCAATAGTTAAGCGGCTTCAAGTCAACATCTGTCGCGTTTTAGCTATGTTCCGG 213311
Qy 361 ACGTCTAGGCAAAAAGTAGTGTGAGATGAAAGCATATCCGTCATTTTAAAGCA 420
Db 213310 ACGTCTAGGCAAAAAGTAGTGTGAGATGAAAGCATATCCGTCATTTTAAAGCA 213252
Qy 421 TCGATAGCTTAATTTGGGCTTAGATCTTCCGCTCTAATAGTATGCAAGACATTCGA 480
Db 213251 TCGATAGCTTAATTTGGGCTTAGATCTTCCGCTCTAATAGTATGCAAGACATTCGA 213192
Qy 481 ATTATTTAACAAGCATTTTTCGGCCGTGAGAAAGCGTTTCCGACTATGTTGGGGC 540
Db 213191 ATTATTTAACAAGCATTTTTCGGCCGTGAGAAAGCGTTTCCGACTATGTTGGGGC 213132
Qy 541 ATGGAACACACTTCAGACTTGAAGCTCATAGACTGGTGTGGAACCTGACAGCTTCAT 600
Db 213131 ATGGAACACACTTCAGACTTGAAGCTCATAGACTGGTGTGGAACCTGACAGCTTCAT 213072
Qy 601 TCTTGGAAATGAATCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660
Db 213071 TCTTGGAAATGAATCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 213012
Qy 661 CGGGCTGAGAGCAAGGCAATGAGATGATGCGTAATTAATGAGAAAGGCAACCTGGAG 720
Db 213011 CGGGCTGAGAGCAAGGCAATGAGATGATGCGTAATTAATGAGAAAGGCAACCTGGAG 212952
Qy 721 GGCATTCGCGTAGCGCTTATTTTGTCCGATTTTCTCTCCGCGGCTTCTTGGGACG 780

|||||
Db 212951 GGCATTCGCGTAGCGCTTATTTTGTCCGATTTTCTCTCTCGCGGCTTCTTGGGACG 212892
Qy 781 GTCCGCTCGTGGCGGATCATGAAAGCGATTCACCGGCCACAGAGCTGAAATCTCCACAG 840
Db 212891 GTCCGCTCGTGGCGGATCATGAAAGCGATTCACCGGCCACAGAGCTGAAATCTCCACAG 212832
Qy 841 CTGGTCTCCCTGCTTCGCGTGGTGGCGGCGCATGACAGAAAGCAATCGACTTTTGTGATG 900
Db 212831 CTGGTCTCCCTGCTTCGCGTGGTGGCGGCGCATGACAGAAAGCAATCGACTTTTGTGATG 212772
Qy 901 ATGGTGTCCATAACCGCGCTGTGACAGCGTCAACCGAGGCCATTTGCCGTTCTGTGATG 960
Db 212771 ATGGTGTCCATAACCGCGCTGTGACAGCGTCAACCGAGGCCATTTGCCGTTCTGTGATG 212712
Qy 961 TATTGGCGAATATCCACAGATGGGTGGCCGATAGGCTGTGGGGTTTCACTGGGACATCTC 1020
Db 212711 TATTGGCGAATATCCACAGATGGGTGGCCGATAGGCTGTGGGGTTTCACTGGGACATCTC 212652
Qy 1021 ACTTTTGGGAAACCGGCGCGCGAGATAGGTTTCTGGGGCTCGGGTGGTGAATTAACC 1080
Db 212651 ACTTTTGGGAAACCGGCGCGCGAGATAGGTTTCTGGGGCTCGGGTGGTGAATTAACC 212592
Qy 1081 ACTGGGCATGCGCTTCCAGACGCTGTGACAGAGCGGAGAAATTTGATGAAAATGAGTGTG 1140
Db 212591 ACTGGGCATGCGCTTCCAGACGCTGTGACAGAGCGGAGAAATTTGATGAAAATGAGTGTG 212532
Qy 1141 ATTGATGGAATTTGTGCGCACTCCAAATTTGCTGCGAGCGGTGGCAAAACCTCAAGATT 1200
Db 212531 ATTGATGGAATTTGTGCGCACTCCAAATTTGCTGCGAGCGGTGGCAAAACCTCAAGATT 212472
Qy 1201 ATTGACCGGATAGAGCAACGGAATGTTTCTCAACAATCTCGGGTGGCACTTCGG 1260
Db 212471 ATTGACCGGATAGAGCAACGGAATGTTTCTCAACAATCTCGGGTGGCACTTCGG 212412
Qy 1261 GTGATGAGAGCGATTCGCGTTCGTGACCGGACGAGAGCTGATCGGGAGATTATG 1320
Db 212411 GTGATGAGAGCGATTCGCGTTCGTGACCGGACGAGAGCTGATCGGGAGATTATG 212352
Qy 1321 GAAACGTTGGGGGAGAGAGCTGCTCAAGCTTTCTGTGTCGCTGGCGGATTAAGCCCG 1380
Db 212351 GAAACGTTGGGGGAGAGAGCTGCTCAAGCTTTCTGTGTCGCTGGCGGATTAAGCCCG 212292
Qy 1381 GCTGTGCGGTTTCCCTGCGCGGCGCATTCGGGGGCGGCGCGTGTGATTTGGGAGAGAT 1440
Db 212291 GCTGTGCGGTTTCCCTGCGCGGCGCATTCGGGGGCGGCGCGTGTGATTTGGGAGAGAT 212232
Qy 1441 CGCGCTTCAAGCTTTGGGCGGACAGAGAGCTGCTTTTGGCGTGTGAGATTTCCGTGCGG 1500
Db 212231 CGCGCTTCAAGCTTTGGGCGGACAGAGAGCTGCTTTTGGCGTGTGAGATTTCCGTGCGG 212172
Qy 1501 CGCGAGCTAAACCTGCGATGCTGTCATTCATTCGACCTTCGCGGCGGAAATGTCGAG 1560
Db 212171 CGCGAGCTAAACCTGCGATGCTGTCATTCATTCGACCTTCGCGGCGGAAATGTCGAG 212112
Qy 1561 GCGGCTGAGAGAGCTGCGATTCGCAAGCTTCATTTGGGCGCACTTTGCAAGCTTATGCAAC 1620
Db 212111 GCGGCTGAGAGAGCTGCGATTCGCAAGCTTCATTTGGGCGCACTTTGCAAGCTTATGCAAC 212052
Qy 1621 GCTCCCTCCCAACCGTTTGGTCAATTAATTTGATGAGGCGTTGGGCGGCGGTGCGC 1680
Db 212051 GCTCCCTCCCAACCGTTTGGTCAATTAATTTGATGAGGCGTTGGGCGGCGGTGCGC 211992
Qy 1681 ATGCTGCCCGCGATCTGCTGTAACCGCGCGCAAAACCGGTGAGCTGTCGCAATGTCACCA 1740
Db 211991 ATGCTGCCCGCGATCTGCTGTAACCGCGCGCAAAACCGGTGAGCTGTCGCAATGTCACCA 211932
Qy 1741 GAGGCGCTGCGGCAATCTCTTCCGACACCAACCAACCGCGGAGAAATGATAGAGCA 1800
Db 211931 GAGGCGCTGCGGCAATCTCTTCCGACACCAACCAACCGCGGAGAAATGATAGAGCA 211872
Qy 1801 CAAGCGTGCAGGCGGACGCACTTTAAAGCAAGGCTTATGACGAGGATGTCGCGGAA 1860

Dh 211871 CAAAGCGTGACGGCGACGACCTTTTAAGCCAAAGGCTTATGACGAGGATGTCGCCGAA 211812
Qy 1861 ACCGAGACTTGTGTAAGAAATTCCTGGGCAATACGACAGCCCTCTCGAATGGAT 1920
Db 211811 ACCGAGACTTGTGTAAGAAATTCCTGGGCAATACGACAGCCCTCTCGAATGGAT 211752
Qy 1921 AACAACTCGGAGAGGCGGCGACGACAGTCGCTTCAACGATTTGAGCGTTCAG 1980
Db 211751 AACAACTCGGAGAGGCGGCGACGACAGTCGCTTCAACGATTTGAGCGTTCAG 211692
Qy 1981 TAAAGAAATTAATGCGCTGATCAATATGATATGAAACCAAGGTTAGCGGCAAGCTG 2040
Db 211691 TAAAGAAATTAATGCGCTGATCAATATGATATGAAACCAAGGTTAGCGGCAAGCTG 211632
Qy 2041 GTGCGCGGAACCTCAGGCGCGGTATGACGCTCTGGGGAATGTCAGCTGACGACGTC 2100
Db 211631 GTGCGCGGAACCTCAGGCGCGGTATGACGCTCTGGGGAATGTCAGCTGACGACGTC 211572
Qy 2101 GCCGACCTTCATGCTGGAATTC 2123
Db 211571 GCCGACCTTCATGCTGGAATTC 211549

RESULT 5
AX127145/c 349980 bp DNA linear PAT 11-MAY-2001
LOCUS AX127145
DEFINITION Sequence 7061 from Patent EP1108790.
ACCESSION AX127145 AX114121
VERSION AX127145.1 GI:14041133
KEYWORDS
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 7061 20-JUN-2001;
JOURNAL KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Source Location/Qualifiers
1. 349980
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
/note="Seq 1 to long (3,309,400) split in 11, seq 7061
0.800.001 0.949.980 ."
BASE COUNT 79274 a 90638 c 98727 g 81341 t
ORIGIN

Query Match 99.4%; Score 2111; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CTCGAGCGGAGAGTGATGATGCGCACCTCTTAAGCAATGCGGCTTTAAATTAAGCAA 60
Db 281620 CTCGAGCGGAGAGTGATGATGCGCACCTCTTAAGCAATGCGGCTTTAAATTAAGCAA 281561
Qy 61 CTTATATGTTTCTACACATCTGCGGACGACGACGAGATATGTTGTCATCAGCTA 120
Db 281560 CTTATATGTTTCTACACATCTGCGGACGACGACGAGATATGTTGTCATCAGCTA 281501
Qy 121 AACGTGGAATGTAAGTACTTAACCTTAACATGCAATGCGATGCGATTTGAAATCTC 180
Db 281500 AACGTGGAATGTAAGTACTTAACCTTAACATGCAATGCGATGCGATTTGAAATCTC 281441
Qy 181 ACTCCCCCAATATCTTAACCTTAACCTTAAGTAAGTATGTTTACCTCATTTATTAAGT 240
Db 281440 ACTCCCCCAATATCTTAACCTTAACCTTAAGTAAGTATGTTTACCTCATTTATTAAGT 281381
Qy 241 TCCCGATCTACCCCTCTTTAACCAGCAATACCCCTTTTGCAAGATTTGCAACACACA 300
Db 281380 TCCCGATCTACCCCTCTTTAACCAGCAATACCCCTTTTGCAAGATTTGCAACACACA 281321

Qy 301 GTGCAATAGTAAACGGGCTTACACCGTACACATTTCTGCGGTTTAAAGGCTATGTTGGG 360
Db 281320 GTGCAATAGTAAACGGGCTTACACCGTACACATTTCTGCGGTTTAAAGGCTATGTTGGG 281261
Qy 361 ACGTCTAGGCAAAAGTAAGTTTGTAGATGAAACGCAATATCCGTCATTTTAAACGCA 420
Db 281260 ACGTCTAGGCAAAAGTAAGTTTGTAGATGAAACGCAATATCCGTCATTTTAAACGCA 281202
Qy 421 TCGATACCTTAATTTGGGCTTAATCTTCCGCTCTTAATAGTATGCAAGACATTGCA 480
Db 281201 TCGATACCTTAATTTGGGCTTAATCTTCCGCTCTTAATAGTATGCAAGACATTGCA 281142
Qy 481 ATTAATTAACAAGGCAATTTTGGCGGAGGAAGGTTTCCGATAGTATGTTGGGCG 540
Db 281141 ATTAATTAACAAGGCAATTTTGGCGGAGGAAGGTTTCCGATAGTATGTTGGGCG 281082
Qy 541 ATGGAACACACTTCAGATTAAGCTCATATGACTCGGTTTGGACCTGACAGCTTCATT 600
Db 281081 ATGGAACACACTTCAGATTAAGCTCATATGACTCGGTTTGGACCTGACAGCTTCATT 281022
Qy 601 TCTTGAATGAATCTCCCAATATGACAACTCATATCAAGGCTATGCAAGACCTTGAG 660
Db 281021 TCTTGAATGAATCTCCCAATATGACAACTCATATCAAGGCTATGCAAGACCTTGAG 280962
Qy 661 CGGCTGGAAGCAAGGCAAAATGAGATGATCGGTATTAATCTGAGAGAGGACCGTGAG 720
Db 280961 CGGCTGGAAGCAAGGCAAAATGAGATGATCGGTATTAATCTGAGAGAGGACCGTGAG 280902
Qy 721 GGCATTCGGGTAGCCGTTATTTTGTCCGATTTTCTCTCTCGCGGTTCTTTGGGACG 780
Db 280901 GGCATTCGGGTAGCCGTTATTTTGTCCGATTTTCTCTCTCGCGGTTCTTTGGGACG 280842
Qy 781 GTGCGCTCGGTGGCATATGAAGGCGATTCACCGGCGCACAGCTGAAATCTCCACTG 840
Db 280841 GTGCGCTCGGTGGCATATGAAGGCGATTCACCGGCGCACAGCTGAAATCTCCACTG 280782
Qy 841 CTGCTCCCTCCCTCTCCGCTGCGGCGGATGCAAGCAATGAGCTTTTGTGATG 900
Db 280781 CTGCTCCCTCCCTCTCCGCTGCGGCGGATGCAAGCAATGAGCTTTTGTGATG 280722
Qy 901 ATGCTGTCATTAACCGGCGCTGTGCAAGCTCAACCGGAGGCGCATTTCCGTTCTG 960
Db 280721 ATGCTGTCATTAACCGGCGCTGTGCAAGCTCAACCGGAGGCGCATTTCCGTTCTG 280662
Qy 961 TATTTGGCAATCCCAAGATGGGTGGCGCATGGCTCTGTGGGTTTATCTGGGCAATCTC 1020
Db 280661 TATTTGGCAATCCCAAGATGGGTGGCGCATGGCTCTGTGGGTTTATCTGGGCAATCTC 280602
Qy 1021 ACTTTGCGGAACCGGCGCGGAGATAGTTTCTGAGGTCCTGCGGTTGAGGATTAAC 1080
Db 280601 ACTTTGCGGAACCGGCGCGGAGATAGTTTCTGAGGTCCTGCGGTTGAGGATTAAC 280542
Qy 1081 ACTGGGATGCGCTTCCAGCGGTGTGCAAGCGGAGCGAGATTTGTGAAATCTG 1140
Db 280541 ACTGGGATGCGCTTCCAGCGGTGTGCAAGCGGAGCGAGATTTGTGAAATCTG 280482
Qy 1141 ATTGATGAATTTGTGTCCCACTTCAATTCGTGCAAGCGGTGCAAAAACCTCAAGTT 1200
Db 280481 ATTGATGAATTTGTGTCCCACTTCAATTCGTGCAAGCGGTGCAAAAACCTCAAGTT 280422
Qy 1201 ATTCAAGCGGTAGAGCAACGATCGTTTCTCCAAACAATCTCTGCGGCGGACCTTCCG 1260
Db 280421 ATTCAAGCGGTAGAGCAACGATCGTTTCTCCAAACAATCTCTGCGGCGGACCTTCCG 280362
Qy 1261 GTGATGAGCGCAATGCGGCTTCTGTGACCCGCAAGGCGCTGGAATCGGGAGATTAAG 1320
Db 280361 GTGATGAGCGCAATGCGGCTTCTGTGACCCGCAAGGCGCTGGAATCGGGAGATTAAG 280302
Qy 1321 GAAACGTTGGGGCAAGCGTGTCAAGCTTTCTGTGCGCGGTCTGGCGCATTTGAGCCG 1380
Db 280301 GAAACGTTGGGGCAAGCGTGTCAAGCTTTCTGTGCGCGGTCTGGCGCATTTGAGCCG 280242

QY 1381 GCTGTGCGGCTTCCCTGCGCGCATCGGGGCGCGCCCGCTGTGCTGATTTGGGAGAT 1440
 Db 280241 GCTGTGCGGCTTCCCTGCGCGCATCGGGGCGCGCCCGCTGTGCTGATTTGGGAGAT 280182
 QY 1441 GCGCGCTTCAAGCTTGGGCGCGAGAGCTGGCTTTGGCGCTGTGGCATTTTGGCTGGG 1500
 Db 280181 GCGCGCTTCAAGCTTGGGCGCGAGAGCTGGCTTTGGCGCTGTGGCATTTTGGCTGGG 280122
 QY 1501 CGCGAGCTTAACTCGCGCATCTGTGCTCATCTGACACCTCGCGCGCGCAATTTGGCGAG 1560
 Db 280121 CGCGAGCTTAACTCGCGCATCTGTGCTCATCTGACACCTCGCGCGCGCAATTTGGCGAG 280062
 QY 1561 GCGGCTGAGAGACTCGCGCATCTGCAAGCTTGGCGCGCACTTTGCAAGCTTATCGAC 1620
 Db 280061 GCGGCTGAGAGACTCGCGCATCTGCAAGCTTGGCGCGCACTTTGCAAGCTTATCGAC 280002
 QY 1621 GCTCCCTTCCCAAGCTTTCGCTCATTTGCTGAGGGCTGGGGCTGGCGCGCTGGCC 1680
 Db 280001 GCTCCCTTCCCAAGCTTTCGCTCATTTGCTGAGGGCTGGGGCTGGCGCGCTGGCC 279942
 QY 1681 ATGCTGCGCGCGCATCTGTGCTCATCTGAGGGCGCGCAAGCGCTGGCTGTGCGCATTTGGCGACA 1740
 Db 279941 ATGCTGCGCGCGCATCTGTGCTCATCTGAGGGCGCGCAAGCGCTGGCTGTGCGCATTTGGCGACA 279882
 QY 1741 GAGGCGGCTTCCGCGCATCTCTTTCGCGCAACCAACAGCGCGGAAATCATAGAGCA 1800
 Db 279881 GAGGCGGCTTCCGCGCATCTCTTTCGCGCAACCAACAGCGCGGAAATCATAGAGCA 279822
 QY 1801 CAGGCGTGTGAGGCGCGCATCTTTTAAAGCAAGGGCTTATGACGGGATCTGTGCGCA 1860
 Db 279821 CAGGCGTGTGAGGCGCGCATCTTTTAAAGCAAGGGCTTATGACGGGATCTGTGCGCA 279762
 QY 1861 ACCGCGCATCTTGTGAGAAATTTCTCGGCAATTCAGAGAGCGCTCTCGCAATTTGAT 1920
 Db 279761 ACCGCGCATCTTGTGAGAAATTTCTCGGCAATTCAGAGAGCGCTCTCGCAATTTGAT 279702
 QY 1921 AACATCCGAGAGGCGCGGAGCGCAAGCTGCTTCAACAGATTTGAGGCTTACGCGAG 1980
 Db 279701 AACATCCGAGAGGCGCGGAGCGCAAGCTGCTTCAACAGATTTGAGGCTTACGCGAG 279642
 QY 1981 TAAAGAAATTTATGCGCTGATCAATTCAGATGTAACACAGGGTACGCGCAGACAGTGG 2040
 Db 279641 TAAAGAAATTTATGCGCTGATCAATTCAGATGTAACACAGGGTACGCGCAGACAGTGG 279582
 QY 2041 GTGCGCGGAACCTCAGGCGCGTATAGCAGCTTCTGGCGGAATGTGACGTGACAGCTTC 2100
 Db 279581 GTGCGCGGAACCTCAGGCGCGTATAGCAGCTTCTGGCGGAATGTGACGTGACAGCTTC 279522
 QY 2101 GCCGACTTTCATGCTGGAATTC 2123
 Db 279521 GCCGACTTTCATGCTGGAATTC 279499
 RESULT 6
 AX066443 1575 bp DNA linear PAT 24-JAN-2001
 LOCUS AX066443
 DEFINITION Sequence 25 from Patent W00100805.
 ACCESSION AX066443
 VERSION AX066443.1 GI:12544151
 KEYWORDS
 SOURCE
 ORGANISM
 Corynebacterium glutamicum
 Corynebacterium glutamicum
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacteriaceae; Corynebacteriidae; Corynebacterium.
 REFERENCE
 1 Pompejus M., Kroeger B., Schroeder H., Zelder O. and Haberman G.
 corynebacterium glutamicum genes encoding proteins involved in
 membrane synthesis and membrane transport
 Patent: WO 0100805-A 25 04-JAN-2001;
 JOURNAL BASF AKTIENGESELLSCHAFT (DE)
 TITL Location/Qualifiers
 FEATURES
 1..1575
 source /organism="Corynebacterium glutamicum"

CDS
 /mol_type="genomic DNA"
 /db_xref="taxon:1718"
 101..1555
 /note="RXA02173"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC26449.1"
 /db_xref="GI:12544152"
 /translation="MVMGMEHTSALTLIDSVLPDPSFISNMTPOYDLNOCVATLE
 PARSRKAKDESVITGEVEGIPVAIVILSDRSFLGSGITVASRIMKILIRATEKL
 PLVSPASGAGRMQEDNRAFWMMVSTIAVOPRHRPAHLPLVLYLNPMTGMAWSGS
 SGHLTFAPRGAQIDGIPRVVLEITGHALPDGQAEMLVTGVIDGIVSPQIAAV
 AKTIKVIQVPEATREFSPPTPGVALPVMEALARSDDPGRPGIEMETIGADVLTSG
 ARAGLSPAVRVALARIGRPVYLIGDPRRTIGPQELRPARGISLARLNPVLSI
 IDTGABLSQAABEHLGASISTARTLSKIDAPLPVYSITIGGVGGGLAMLPADLVY
 ALENMUSALPPEGASALIFPDTHNMASTIERQGVQALALLSQGLIDIVARTEHVE
 EILGTSNALSLENNPERRARDSPFTFERLIAQ"
 BASE COUNT 309 a 434 c 474 g 358 t
 ORIGIN
 Query Match 74.2%; Score 1575; DB 6; Length 1575;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 429 CTAATTGGGCTTAGATCTTCGCGCTTAATAGGTATGAGAGCATTCGAAATTAATTA 488
 Db 1 CTAATTGGGCTTAGATCTTCGCGCTTAATAGGTATGAGAGCATTCGAAATTAATTA 60
 QY 489 ACAAGCCATTTTTCGCGCGGTGAGAGAGCGTTCCTGACTATGATGTGGGGCATGAGACA 548
 Db 61 ACAAGCCATTTTTCGCGCGGTGAGAGAGCGTTCCTGACTATGATGTGGGGCATGAGACA 120
 QY 549 CACTTACAGATTGACGCTCATAGACTCGGTTTGGACCTGACAGCTTCAATTTCTTGA 608
 Db 121 CACTTACAGATTGACGCTCATAGACTCGGTTTGGACCTGACAGCTTCAATTTCTTGA 180
 QY 609 TGAATCTCCCAATATGACCACTTCATCAAGGCTATGACAGACCTTGGAGCGGGCTCG 668
 Db 121 TGAATCTCCCAATATGACCACTTCATCAAGGCTATGACAGACCTTGGAGCGGGCTCG 240
 QY 669 AAGCAAGCCAAATGCGATGTAATCGTAAATTAAGTGAAGAGCAACCGTGAGGGCATTC 728
 Db 241 AAGCAAGCCAAATGCGATGTAATCGTAAATTAAGTGAAGAGCAACCGTGAGGGCATTC 300
 QY 729 GGTAGCGCTTATTTTGTCCGATTTTCTCTCGCGCGCTTCTTGGGACGCTGCGCTC 788
 Db 301 GGTAGCGCTTATTTTGTCCGATTTTCTCTCGCGCGCTTCTTGGGACGCTGCGCTC 360
 QY 789 GGTGGGATCATGAAAGGATTCACCGGCGCAAGAGCTGAAATCTCCACGCTGTGCTC 848
 Db 361 GGTGGGATCATGAAAGGATTCACCGGCGCAAGAGCTGAAATCTCCACGCTGTGCTC 420
 QY 849 CCTGCTTCCGCTGCTGCGCGCATGACAGAGAAATGAGCTTTTGTGATGATGATGTC 908
 Db 421 CCTGCTTCCGCTGCTGCGCGCATGACAGAGAAATGAGCTTTTGTGATGATGATGTC 480
 QY 909 CATTAACCGGCTGTGACGCTGACCGGAGCGCATTTGCGCTTCTGTGTATTTGGCG 968
 Db 481 CATTAACCGGCTGTGACGCTGACCGGAGCGCATTTGCGCTTCTGTGTATTTGGCG 540
 QY 969 CAATCCACGATGGGTGGCGCATAGCGCTCTGTGGGGTTTCAATGAGGATCTCACTTTTGC 1028
 Db 541 CAATCCACGATGGGTGGCGCATAGCGCTCTGTGGGGTTTCAATGAGGATCTCACTTTTGC 600
 QY 1029 GGAACCGGCGCGGAGATAGATTCTTGGGCTCTGCGGTGTGAGTTAACTGAGGCA 1088
 Db 601 GGAACCGGCGCGGAGATAGATTCTTGGGCTCTGCGGTGTGAGTTAACTGAGGCA 660
 QY 1089 TGGCCTTCCAGACGCTGTGACAGCGCGGAGAAATTTGTGAAAATCTGTGTGATGATGG 1148
 Db 661 TGGCCTTCCAGACGCTGTGACAGCGCGGAGAAATTTGTGAAAATCTGTGTGATGATGG 720
 QY 1149 AATTGTGTGCGCATCTCCAAATTTGCTGTGACGGGTGGCAAAACCTTCAAGTTATTCAGCC 1208

```

|||||
Db 721 AATGTGTCCGCACTCCAAATTGGCGAGCGGTGGCAAAAACCTCAAGATTATTCAGCC 780
|||
Qy 1209 GGTGAGGCAACGGATCGTTTTTCTCCACAACACTCTGGGTGGCACTTCGGGTGATGGA 1268
|||
Db 781 GGTGAGGCAACGGATCGTTTTTCTCCACAACACTCTGGGTGGCACTTCGGGTGATGGA 840
|||
Qy 1269 GGGGATTTGGCGGCTTCTGTGACCCGCGAGAGGCTGGAAATGGGGGATTTATGAAAGCTT 1328
|||
Db 841 GGGGATTTGGCGGCTTCTGTGACCCGCGAGAGGCTGGAAATGGGGGATTTATGAAAGCTT 900
|||
Qy 1329 GGGGGCAGACGTCGTCAAGCTTTCTGTGCGCGTGTGCGGCATTTGAGCCCGGCTGTGCG 1388
|||
Db 901 GGGGGCAGACGTCGTCAAGCTTTCTGTGCGCGTGTGCGGCATTTGAGCCCGGCTGTGCG 960
|||
Qy 1389 CGTTGCCCTTGCGCGCATCGGGGGCCGGCCCGTGTGCTGATTGGGAGAGATCGCGCTT 1448
|||
Db 961 CGTTGCCCTTGCGCGCATCGGGGGCCGGCCCGTGTGCTGATTGGGAGAGATCGCGCTT 1020
|||
Qy 1449 CAGGCTTGGGCGCGAGAGAGCTGGCTTGGCGGTGCGGCACTTTCGCTGGGCGCGAGCT 1508
|||
Db 1021 CAGGCTTGGGCGCGAGAGAGCTGGCTTGGCGGTGCGGCACTTTCGCTGGGCGCGAGCT 1080
|||
Qy 1509 AAACCTGCCGATCGTGTTCATCATCAACCTCCGCGCGCGAATTGTGCGAGCGGCTGA 1568
|||
Db 1081 AAACCTGCCGATCGTGTTCATCATCAACCTCCGCGCGCGAATTGTGCGAGCGGCTGA 1140
|||
Qy 1569 GAGAGCTGGCGCATCGCAAGCTCGATTGGCGCGCACTTGTCAAGCTTATCGACTCCCT 1628
|||
Db 1141 GAGAGCTGGCGCATCGCAAGCTCGATTGGCGCGCACTTGTCAAGCTTATCGACTCCCT 1200
|||
Qy 1629 CCCGACGCTTTCGATCATTTATTGGTCAAGGGGCTTGGCGGTCGCGCGCATGCTGCC 1688
|||
Db 1201 CCCGACGCTTTCGATCATTTATTGGTCAAGGGGCTTGGCGGTCGCGCGCATGCTGCC 1260
|||
Qy 1689 CGCGGATCTGTCTTACGCGCGCGCAAAACGCGTGTGCTGTCCGATTGCCACAGAGGCGC 1748
|||
Db 1261 CGCGGATCTGTCTTACGCGCGCGCAAAACGCGTGTGCTGTCCGATTGCCACAGAGGCGC 1320
|||
Qy 1749 CTCGGGCAATCTCTTCGCGCAACCCAAACACGCGCGGAAATCATAGAGCAAAAGCGT 1808
|||
Db 1321 CTCGGGCAATCTCTTCGCGCAACCCAAACACGCGCGGAAATCATAGAGCAAAAGCGT 1380
|||
Qy 1809 GAGAGCGAGCACTTTTAAGCAAGGCTTATGAGGAGTCTGCGCGCAAAACGAGCA 1868
|||
Db 1381 GAGAGCGAGCACTTTTAAGCAAGGCTTATGAGGAGTCTGCGCGCAAAACGAGCA 1440
|||
Qy 1869 CTTTGTGAAGAAATTCGCGCAATAGCAAGCGCTCTCCGAATTGATTAACAATCC 1928
|||
Db 1441 CTTTGTGAAGAAATTCGCGCAATAGCAAGCGCTCTCCGAATTGATTAACAATCC 1500
|||
Qy 1929 GAGAGGCGGAGCGAGCAAGTCTGCTTCAACGATTGAGCGTTAGCGCAAGTAAAGAA 1988
|||
Db 1501 GAGAGGCGGAGCGAGCAAGTCTGCTTCAACGATTGAGCGTTAGCGCAAGTAAAGAA 1560
|||
Qy 1989 ATTATGCGCTGATCA 2003
|||
Db 1561 ATTATGCGCTGATCA 1575
|||

```

```

RESULT 7
AR202320 1473 bp DNA linear PAT 20-APR-2002
LOCUS AR202320
DEFINITION Sequence 2 from patent US 6361986.
ACCESSION AR202320
VERSION AR202320.1 GI:20256859
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1473)
AUTHORS Tilly, Y., Eikmanns, B., Eggeling, L., Sahn, H. and Mockel, B.
TITLE Process for the preparation of L-amino acids by fermentation and

```

```

JOURNAL nucleotide sequences coding for the accD gene
Patent: US 6361986-A 2 26-MAR-2002;
FEATURES
SOURCE Location/Qualifiers
1..1473
BASE COUNT 275 a 414 c 456 g 328 t
ORIGIN
Query Match 69.4%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 508 GTGAGAGAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 567
Db 1 GTGAGAGAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
Qy 568 ATAGACTCGGTTTGAACCCCTGACAGCTTCAATTTCTTGAATGAATCCCAATATGAC 627
Db 61 ATAGACTCGGTTTGAACCCCTGACAGCTTCAATTTCTTGAATGAATCCCAATATGAC 120
Qy 628 AACCTCAATCAAGGCTATGAGAGACCTTGGACCGGCTCGAAGCAAGGCAATGGAT 687
Db 121 AACCTCAATCAAGGCTATGAGAGACCTTGGACCGGCTCGAAGCAAGGCAATGGAT 180
Qy 688 GAATCGGTAATTACTGGAGAGAGCACCGTGAAGGCAATTCGGTAAAGCGTTATTTGTC 747
Db 181 GAATCGGTAATTACTGGAGAGAGCACCGTGAAGGCAATTCGGTAAAGCGTTATTTGTC 240
Qy 748 GATTTTCTTCTCGCGGCTTCTTTGGACAGCGTGTGCGATCATGAAGCG 807
Db 241 GATTTTCTTCTCGCGGCTTCTTTGGACAGCGTGTGCGATCATGAAGCG 300
Qy 808 ATTACCGGCGCACAGAGCTGAAACCTCCACTGTGTCTCCCTGCTCGGTGGTGGC 867
Db 301 ATTACCGGCGCACAGAGCTGAAACCTCCACTGTGTCTCCCTGCTCGGTGGTGGC 360
Qy 868 CGCATGAGAGAAACAAATCAGCTTTTGTATGATGTGTCCATTAACCGCGGCTGTGAC 927
Db 361 CGCATGAGAGAAACAAATCAGCTTTTGTATGATGTGTGTCCATTAACCGCGGCTGTGAC 420
Qy 928 CGTACCGGAGAGCGCAATTTGCCGTTCTGTGTATTTGGCAATCCAGATGGGTGC 987
Db 421 CGTACCGGAGAGCGCAATTTGCCGTTCTGTGTATTTGGCAATCCAGATGGGTGC 480
Qy 988 GCCATGGCTCTGTGGGTTTATCTGGGCACTCACTTTTGGGAACCCGCGGCGCAGATA 1047
Db 481 GCCATGGCTCTGTGGGTTTATCTGGGCACTCACTTTTGGGAACCCGCGGCGCAGATA 540
Qy 1048 GGTTCCTGGGTCCTCGCGTGTGTGAGTTAACCACTGGGCGATGCGCTTCAGACGCTGTG 1107
Db 541 GGTTCCTGGGTCCTCGCGTGTGTGAGTTAACCACTGGGCGATGCGCTTCAGACGCTGTG 600
Qy 1108 CAGCAGCGGAGAAATTTGTAAGAACTGTGTGATGAGAAATTTGTGCGCACTTCAA 1167
Db 601 CAGCAGCGGAGAAATTTGTAAGAACTGTGTGATGAGAAATTTGTGCGCACTTCAA 660
Qy 1168 TTGCTGTACAGCGTGTGAAAAACCTTCAAGTTATTCAGCCGTTAGAGCAAGATCGT 1227
Db 661 TTGCTGTACAGCGTGTGAAAAACCTTCAAGTTATTCAGCCGTTAGAGCAAGATCGT 720
Qy 1228 TTTTCTCAACCACTCTGCGGTGTGGAACCTTCGAGTGTGAGGCGATTGCGCTTCGT 1287
Db 721 TTTTCTCAACCACTCTGCGGTGTGGAACCTTCGAGTGTGAGGCGATTGCGCTTCGT 780
Qy 1288 GACCCGAGAGGCTGTGAATCGGAGAGATTATGAAACGTTGGGCGAGACGTCGTAAG 1347
Db 781 GACCCGAGAGGCTGTGAATCGGAGAGATTATGAAACGTTGGGCGAGACGTCGTAAG 840
Qy 1348 CTTTCTGTGTGCGGTGTGCGCATTTGAGCCGCTGTGTGCGCTTGTGCGCGCATC 1407
Db 841 CTTTCTGTGTGCGGTGTGCGCATTTGAGCCGCTGTGTGCGCTTGTGCGCGCATC 900
Qy 1408 GGGGGCGGCGCGGTGTGTGATTTGGGCAAGATCGCGCTTCAAGCTGTGGCGCGAGAG 1467

```

```

Db      901 GGGGGCCGGCCGCTGCTGATGGGAGATCGCGCTTCACGCTTGAGCGCGAGAG 960
Qy      1468 CTGCGTTTGGCGCTGCTGATGGGAGATCGCGCTTCACGCTTGAGCGCGAGAG 1527
Db      961 CTGCGTTTGGCGCTGCTGATGGGAGATCGCGCTTCACGCTTGAGCGCGAGAG 1020
Qy      1528 ATCATCGACACCTCGCGCGCGAGATGGTGGAGCGCGCTGAGAGAGCTGGAGATCGCAAGC 1587
Db      1021 ATCATCGACACCTCGCGCGCGAGATGGTGGAGCGCGCTGAGAGAGCTGGAGATCGCAAGC 1080
Qy      1588 TCGATTGGCGCGACCTTGTCCAGCTTATCGACGCTCCCTCCCGACCGCTTTCGGTCAAT 1647
Db      1081 TCGATTGGCGCGACCTTGTCCAGCTTATCGACGCTCCCTCCCGACCGCTTTCGGTCAAT 1140
Qy      1648 ATTGGTCAAGGGCGTTGGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 1707
Db      1141 ATTGGTCAAGGGCGTTGGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 1200
Qy      1708 GCCGAAAACGCGTGGCTGTCCGCAATTCGACGAGAGAGCGCGCTGGCGCTGGCGCTGGCGCT 1767
Db      1201 GCCGAAAACGCGTGGCTGTCCGCAATTCGACGAGAGAGCGCGCTGGCGCTGGCGCTGGCGCT 1260
Qy      1768 GACACCAACGACGCGCGGAAATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1827
Db      1261 GACACCAACGACGCGCGGAAATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy      1828 AGCCAAAGGCTTATGACGCGGATTCGTCGCGGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1887
Db      1321 AGCCAAAGGCTTATGACGCGGATTCGTCGCGGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1888 GGCACATCAGCAAGCGCTCTCCGAAATTCGATTAACAATCGGAGAGAGAGAGAGAGAGAGAGAG 1947
Db      1381 GGCACATCAGCAAGCGCTCTCTCCGAAATTCGATTAACAATCGGAGAGAGAGAGAGAGAGAGAG 1440
Qy      1948 AGTCGCTTCAACAGATTGAGCGCTTAAAGCGCAG 1980
Db      1441 AGTCGCTTCAACAGATTGAGCGCTTAAAGCGCAG 1473

```

```

RESULT 8
AX121013      1473 bp      DNA      linear      PAT 11-MAY-2001
LOCUS      AX121013
DEFINITION      Sequence 929 from Patent EP1108790.
ACCESSION      AX121013
VERSION      AX121013.1 GI:14037728
KEYWORDS
SOURCE      Corynebacterium glutamicum
              Corynebacterium glutamicum
              Bacteria; Actinobacteria; Actinomycetales;
              Corynebacteriaceae; Corynebacterium.
REFERENCE
  1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
    Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
    Novel polymucleotides
    Patent: EP 1108790-A 929 20-JUN-2001;
    JOURNAL      KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
  source
    1. 1473
    /organism="Corynebacterium glutamicum"
    /mol_type="genomic DNA"
    /db_xref="taxon:1718"
BASE COUNT      275 a      414 c      456 g      328 t
ORIGIN

```

```

Query Match      69.4%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      508 GTGGAAGCGTTTCCGACTATGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 567
Db      1 GTGGAAGCGTTTCCGACTATGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60

```

```

Qy      568 ATAGACGCGTTTGGAGACCTTGACAGCTTCAATTTCTGGAAATGAATCTCCCAATATGAC 627
Db      61 ATAGACGCGTTTGGAGACCTTGACAGCTTCAATTTCTGGAAATGAATCTCCCAATATGAC 120
Qy      628 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 687
Db      121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy      688 GAAATGGTAATTAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
Db      181 GAAATGGTAATTAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy      748 GATTTTCTTCTCTGCGGCTTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
Db      241 GATTTTCTTCTCTGCGGCTTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      808 ATTCAAGCGCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867
Db      301 ATTCAAGCGCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy      868 CGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927
Db      361 CGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      928 CGTCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987
Db      421 CGTCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      988 GGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047
Db      481 GGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      1048 GGTTCCTGCGGCTCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1107
Db      541 GGTTCCTGCGGCTCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      1108 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1167
Db      601 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      1168 TTGCGTGAAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1227
Db      661 TTGCGTGAAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy      1228 TTTTCTCCAACTCTCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1287
Db      721 TTTTCTCCAACTCTCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy      1288 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1347
Db      781 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy      1348 CTTTCTGAGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407
Db      841 CTTTCTGAGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy      1408 GGGGGGCGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1467
Db      901 GGGGGGCGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy      1468 CTGCGTTTGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527
Db      961 CTGCGTTTGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy      1528 ATCATCGACACCTCGCGCGCGAGATGGTGGAGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1587
Db      1021 ATCATCGACACCTCGCGCGCGAGATGGTGGAGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy      1588 TCGATTGGCGCGACCTTGTCCAGCTTATCGACGCTCCCTCCCGACCGCTTTCGGTCAAT 1647
Db      1081 TCGATTGGCGCGACCTTGTCCAGCTTATCGACGCTCCCTCCCGACCGCTTTCGGTCAAT 1140
Qy      1648 ATTGGTCAAGGGCGTTGGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCT 1707

```


Db	1141	ATTGGTCAGGGCGCTTGCGCGTGGCGCGCTTGGCATGCTGCGCCGCGCATCTGTCTACGCG	1200
QY	1708	GCCGAAAACGCGTGGCTGTCCGCATTTGCGACACAGAGGGCGCTCGGCCATCTCTTCGCG	1767
Db	1201	GCCGAAAACGCGTGGCTGTCCGCATTTGCGACACAGAGGGCGCGCTCGGCCATCTCTTCGCG	1260
QY	1768	GACACCAACCAACGCGCGCGGAAATCATTAAGGAGCAAGGCGTGGAGCGGCACAGCACTTTTA	1827
Db	1261	GACACCAACCAACGCGCGCGGAAATCATTAAGGAGCAAGGCGTGGAGCGGCACAGCACTTTTA	1320
QY	1828	AGCCAGAGGGCTTATGACGGGATCGTCCCGCAAAACCGAGCATTTGTGTGAAGAAATTCTC	1887
Db	1321	AGCCAGAGGGCTTATGACGGGATCGTCCCGCAAAACCGAGCATTTGTGTGAAGAAATTCTC	1380
QY	1888	GGCACATATCAGCAACGCCCTCTCCGAATTGGATTAACAATTCGGAGAGGGCGGACGCGAC	1947
Db	1381	GGCACATATCAGCAACGCCCTCTCCGAATTGGATTAACAATTCGGAGAGGGCGGACGCGAC	1440
QY	1948	AGTCGCTTACACAGATTTGAGGGTTAGCGGAG	1980
Db	1441	AGTCGCTTACACAGATTTGAGGGTTAGCGGAG	1473

LOCUS	DEFINITION	1473 bp	DNA	linear	PAT 31-JAN-2002
BD004742	Replicable DNA, amino acid sequence, Corynebacterium microorganism.				
BD004742	shuttle vector, and process for producing L-amino acid.				
BD004742.1	GI:18632703				
JP 2001008693-A/2.					
ORGANISM	Corynebacterium glutamicum				
REFERENCE	Corynebacterium glutamicum				
AUTHORS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.				
TITLE	1 (bases 1 to 1473)				
JOURNAL	Tilke, I., Eggering, L., Eickmans, B., Zamu, H. and Meck, V.				
COMMENT	shuttle vector, and process for producing L-amino acid				
DECUSSA HUELS AG, FORSCHUNGSZENTRUM JUELICH GMBH	Patent: JP 2001008693-A 2 16-JAN-2001;				
OS	Corynebacterium glutamicum				
PN	JP 2001008693-A/2				
PD	16-JAN-2001				
PF	24-MAY-2000 JP 2000153547				
PR	27-MAY-1999 DE 19924365.4				
PI	IVENNE TILKE, LOTHAR EGGERING, BERNHARD EICKMANS, HERMANN ZAMU, PI VETTINA MECKEL				
PC	C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, PC C12P13/08,				
PC	C12P13/12, C12P13/20, C12P21/02/(C12N15/09, C12R1:15), (C12N1/21,				
PC	C12R1:15),				
PC	(C12P13/04, C12R1:15), (C12P13/06, C12R1:15), (C12P13/08, C12R1:15), PC				
PC	C12N15/12, C12R1:15), (C12P13/20, C12R1:15), (C12P21/02, C12R1:15), PC				
PC	C12N15/00, C12R1:15)				
CC					
CC					
FEATURES					
source	Location/Qualifiers				
FT	CDS				
Key	Location/Qualifiers				
1..1473					
/organism="Corynebacterium glutamicum"					
/mol_type="genomic DNA"					
/db_xref="taxon:1718"					
BASE COUNT	275 a 414 c 456 g 328 t				
ORIGIN					
Query Match	69.4%; Score 1473; DB 6; Length 1473;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 1473; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				

OY	508	GTGAGAAAGGTTTTCCGACTAATGAGTGGGGCATGGAAACAACCTTACGATTTGAGCTC	567
Db	1	GTGGAGAAGGTTTTCCGACTAATGAGTGGGGCATGGAAACAACCTTACGATTTGAGCTC	60
OY	568	ATAACTCGTTTTGGAACCTTGCACGCTTCAATTTCTTGAATGAAACTCCCAATATGAC	627
Db	61	ATAACTCGTTTTGGAACCTTGCACGCTTCAATTTCTTGAATGAAACTCCCAATATGAC	120
OY	628	AACTCAATCAAGGCTTTGCAAGACCTTTGAGCGGGCTGGAGCAAGGCCAAATGCGAT	687
Db	121	AACTCAATCAAGGCTTTGCAAGACCTTTGAGCGGGCTGGAGCAAGGCCAAATGCGAT	180
OY	688	GAATCGGTAATTAATGAGAAAGGACCGTGAAGGGCAATTCGGATAGCGCTTAATTTGTC	747
Db	181	GAATCGGTAATTAATGAGAAAGGACCGTGAAGGGCAATTCGGATAGCGCTTAATTTGTC	240
OY	748	GATTTTCTCTTCCCTGGCGGTTCTTTTGGGACGGTCCCGTGGTGCAGATCATGAAAGCG	807
Db	241	GATTTTCTCTTCCCTGGCGGTTCTTTTGGGACGGTCCCGTGGTGCAGATCATGAAAGCG	300
OY	808	ATTACCGCGCCACAGAGCTGAAACTCCACCTGCTGGTCTCCCTGCTTCGGTGGTGGC	867
Db	301	ATTACCGCGCCACAGAGCTGAAACTCCACCTGCTGGTCTCCCTGCTTCGGTGGTGGC	360
OY	868	CGCATGCAAGAAACATTCAGCTTTTGATATGATGATGATGATCAACCGGCGCTGTCAG	927
Db	361	CGCATGCAAGAAACATTCAGCTTTTGATATGATGATGATGATCAACCGGCGCTGTCAG	420
OY	928	CGTACCGCGAGGCGCATTTGCGCTTCTGGTGTATTTGGCGAATCCACGATGGTGGC	987
Db	421	CGTACCGCGAGGCGCATTTGCGCTTCTGGTGTATTTGGCGAATCCACGATGGTGGC	480
OY	988	GCCATGGCTCTGTTGGGTTATCTGGGCATCTCACTTTTGGGAAACCGGCGCGCAATA	1047
Db	481	GCCATGGCTCTGTTGGGTTATCTGGGCATCTCACTTTTGGGAAACCGGCGCGCAATA	540
OY	1048	GGTTTCTCGGTCTCTCCCGTGTGGGAATTAAACACTGGGCAATGGCGTTCCAGAGGTGTG	1107
Db	541	GGTTTCTCGGTCTCTCCCGTGTGGGAATTAAACACTGGGCAATGGCGTTCCAGAGGTGTG	600
OY	1108	CAGCAGGCGAGATTTTGTGCAAAACTGGTGTATGATGAAATTTGTGCGCACTCCAA	1167
Db	601	CAGCAGGCGAGATTTTGTGTAANAACGTGTGTATTAATGAAATTTGTGCGCACTCCAA	660
OY	1168	TTGGGTGACGGGTGGCAAAAACCTCAAGGTATTTACGCCGTAGAGGCAAGGATCGT	1227
Db	661	TTGGGTGACGGGTGGCAAAAACCTCAAGGTATTTACGCCGTAGAGGCAAGGATCGT	720
OY	1228	TTTTTCTCCAAACTCTCTGGGTGGCACTTCGGGTATGAGGGGATTTGCGCTTCGT	1287
Db	721	TTTTTCTCCAAACTCTCTGGGTGGCACTTCGGGTATGAGGGGATTTGCGCTTCGT	780
OY	1288	GACCCGAGAGGCTTGGATTCGSGGAGATTAATGAAAAGTTGGGGGAGACGTCGTCAAG	1347
Db	781	GACCCGAGAGGCTTGGATTCGSGGAGATTAATGAAAAGTTGGGGGAGACGTCGTCAAG	840
OY	1348	CTTTCTGGTGGCGTGTCTGGCAATTAGACCCCGCTGTGCGCGTTGCTGCTGGCGCATC	1407
Db	841	CTTTCTGGTGGCGTGTCTGGCAATTAGACCCCGCTGTGCGCGTTGCTGCTGGCGCATC	900
OY	1408	GGGGGCGGGCCGTGTGTCTGATTTGGGCAAGATCGCGCTTACACGTTTGGGCGCAGAG	1467
Db	901	GGGGGCGGGCCGTGTGTCTGATTTGGGCAAGATCGCGCTTACACGTTTGGGCGCAGAG	960
OY	1468	CTGGCTTTTGGGCGTCGTGGCATTTCCGTGGCGGGCAGAGCTAAACCTTGCATGTGTCC	1527
Db	961	CTGGCTTTTGGGCGTCGTGGCATTTCCGTGGCGGGCAGAGCTAAACCTTGCATGTGTCC	1020
OY	1528	ATCATTCAGACCTTCGGGCGCGAATTTGCGAGCGGCTGAGGACTTCGAGATCGCAAGC	1587
Db	1021	ATCATTCAGACCTTCGGGCGCGAATTTGCGAGCGGCTGAGGACTTCGAGATCGCAAGC	1080
OY	1588	TGATATGGCGACCTTGTCCAACTTAATGACGCTCCCTCCCAACCGTTTCGATATT	1647

Db	1081	TCGATTGGCGGCACCTTGTGCCAGTTTACAGCGCTCCCTTCCACCGGTTGGGTCA	1140
Qy	1648	ATTGTCAAGGCGGTGGCGGTGGCGCGCTGGCCATGTCCCGCCGATCTGGTCTACCG	1707
Db	1141	ATTGTCAAGGGGTTGGCGGTGGCGCGCTGGCGCATGTCTCCCGCGATCTGATCTACCG	1207
Qy	1708	GCCGAAAACGGGTGGCTGTCCGCATTTGCCACCAAGAGGGCGCTCGGCAATCTCTTCGG	1767
Db	1201	GCCGAAAACGGGTGGCTGTCCGCATTTGCCACCAAGAGGGCGCTCGGCAATCTCTTCGG	1267
Qy	1768	GACACCAACCAAGCGCCGCGGAAATCATAGACGCAAGGCGTGGCAGCGCAACGCACTTTTA	1827
Db	1261	GACACCAACCAAGCGCCGCGGAAATCATAGACGCAAGGCGTGGCAGCGCAACGCACTTTTA	1327
Qy	1828	AGCCCAAGGGCTTATGACGGGATGTGTCGCCAAACGAGCACTTTGTTGAAGAAATTTCT	1887
Db	1321	AGCCCAAGGGCTTATGACGGGATGTGTCGCCAAACGAGCACTTTGTTGAAGAAATTTCT	1387
Qy	1888	GGCACAATCAGCAACGCCCTCTCCGAATTGATTAACAATCCGGAGAGGGCGGAGCGCAC	1947
Db	1381	GGCACAATCAGCAACGCCCTCTCCGAATTGATTAACAATCCGGAGAGGGCGGAGCGCAC	1447
Qy	1948	AGTGGCTTACACAGATTGAGCGTTTACGCGAG	1980
Db	1441	AGTGGCTTACACAGATTGAGCGTTTACGCGAG	1473

RESULT 10			
BD163130			
LOCUS	BD163130	1473 bp	DNA
DEFINITION	Novel polynucleotide.		linear
ACCESSION	BD163130		
VERSION	BD163130.1 GI:27868892		
KEYWORDS	JP 2002191370-A/929.		
SOURCE	unidentified		
			PAT 17-JAN-2003

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1473)	Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ohishi, K., Yocoi, H., Tateishi, N., Senoo, A., Ikeda, H., and Ozaki, A.	Novel polynucleotide	Patent: JP 2002191370-A 929 09-JUL-2002;	
	KIOWA HAKKO KOGYO CO LTD	<i>Corynebacterium glutamicum</i>		

Accession	Source	Location/Qualifiers
PF 15-DEC-2000 JP 2000405096		
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,		
PI KEIKO OCHIAI,		
PI HARUHIKO YOKOI, NAOKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO		
PI OZAKI		
PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC		
C12M1/15,		
PC C12M1/19, C12M1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/		
PC 04, C12P13/08,		
PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC		
G01N33/566,		
PC G01N33/569, G01N33/68, G01N37/00//C12P21/08, (C12M1/21, C12R1:15),		
PC (C12M1/21, C12R1:13), (C12M1/21, C12R1:01), (C12P13/08, C12R1:15),		
PC C12N15/00,		
PC C12N5/00, C12N15/00		
CC Novel polynucleotide		
PH Key		
FT source	1.1473	
	/organism='Corynebacterium glutamicum'.	

FEATURES	source	location/qualifiers
BASE COUNT	275 a	1..1473 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"
ORIGIN	414 c	456 g 328 t

Query Match	69.4%	Score 1473;	DB 6;	Length 1473;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1473; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	508	TTGGAGGAGCGCTTTTCCACATATGGTGTGGGGCATATGAAACAACATTCACGATTTAGCGCTC	567
Db	1	GTGGAGGAAGCGTTTTCCGATATGGTGTGGGGCATATGAAACAACATTCACGATTTAGCGCTC	60
QY	558	ATGAGCTCGGTTTTGGACCCTGACAGCTTCATTTCTTGGATGAATCCGCCAATATGAC	627
Db	61	ATGAGCTCGGTTTTGGACCCTGACAGCTTCATTTCTTGGATGAATCCGCCAATATGAC	120
QY	628	AACCGCATCAAGCGTATGCAAGACCTTTGGAGCGGGCTCCAAAGAAAGCCAAATGCCAT	687
Db	121	AACCGCATCAAGCGTATGCAAGACCTTTGGAGCGGGCTCCAAAGAAAGCCAAATGCCAT	180
QY	688	GAATCGGTAAATTACTGAGAAAGGACCGTGAAGGGCATTCGCGTATGCGTTATTTGTCC	747
Db	181	GAATCGGTAAATTACTGAGAAAGGACCGTGAAGGGCATTCGCGTATGCGTTATTTGTCC	240
QY	748	GATTTTTCCTTCCTCGGCGGCTTCTTTGGGCAAGGTCGCTCGCTGCTGCAATCATGAAGCG	807
Db	241	GATTTTTCCTTCCTCGGCGGCTTCTTTGGGCAAGGTCGCTCGCTGCTGCAATCATGAAGCG	300
QY	808	ATTACCGCGGCAAGAGCTGAATCCCACTGCTGCTCCCTCGCTTCGCGTGGTTCG	867
Db	301	ATTACCGCGGCAAGAGCTGAATCCCACTGCTGCTCCCTCGCTTCGCGTGGTTCG	360
QY	868	CGCATGCAAGAAACAATCGAGCTTTTGTTCATGATGAGTGCATTAACCGCGGCTTGGAG	927
Db	361	CGCATGCAAGAAACAATCGAGCTTTTGTTCATGATGAGTGCATTAACCGCGGCTTGGAG	420
QY	928	CGTACCGCGGAGCGGCATTTGCGGTTCTCGGTGTAATTTGGCAATCCACGATGGGTGC	987
Db	421	CGTACCGCGGAGCGGCATTTGCGGTTCTCGGTGTAATTTGGCAATCCACGATGGGTGC	480
QY	988	GCCATGGCGCTGTGAGGGGTTTCATCTGAGGAGATCTCATTTTGGGAAACCGGCGGCGAGATA	1047
Db	481	GCCATGGCGCTGTGAGGGGTTTCATCTGAGGAGATCTCATTTTGGGAAACCGGCGGCGAGATA	540
QY	1048	GGTTTCCGTGGTCCCTCGCGTGTGGAGTTAAACAATGGGCACTGGCGCTTCCAGACGATGTG	1107
Db	541	GGTTTCCGTGGTCCCTCGCGTGTGGAGTTAAACAATGGGCACTGGCGCTTCCAGACGATGTG	600
QY	1108	CAGCAGCGCGAAGAAATTTGTTGTAACACTGTGTGTAATTGATGGAATTTGTGTGCCACTCCAA	1167
Db	601	CAGCAGCGCGGAGAAATTTGTTGTAACACTGTGTGTAATTGATGGAATTTGTGTGCCACTCCAA	660
QY	1168	TTGGGTGAGCGGTGGGAAAAACCTCAAGGTTATTACCGCGTATGAGGCAACGGAATGCT	1227
Db	661	TTGGGTGAGCGGTGGGAAAAACCTCAAGGTTATTACCGCGTATGAGGCAACGGAATGCT	720
QY	1228	TTTTCTCAACAACATCTCCGCGTGTGGCACTTCGCGTGAATGAGAGCGAATTCGCGTTCTCGT	1287
Db	721	TTTTCTCAACAACATCTCCGCGTGTGGCACTTCGCGTGAATGAGAGCGAATTCGCGTTCTCGT	780
QY	1288	GACCCGCAAGAGCCTTGGAAATCGGGGAGATTATGAAACGTTGGGGGCAACGTCGTCAAG	1347
Db	781	GACCCGCAAGAGCCTTGGAAATCGGGGAGATTATGAAACGTTGGGGGCAACGTCGTCAAG	840
QY	1348	CTTTCCTGAGCGGTGCTGGGCGCAATTAGGCCCGGCTGTGGCGCTTTGGCGCGCGCATC	1407
Db	841	CTTTCCTGAGCGGTGCTGGGCGCAATTAGGCCCGGCTGTGGCGCTTTGGCGCGCGCATC	900
QY	1408	GGGGGCGGCGCCGTGATGCTGATTTGGGCAAGATCGCGCTTCAACGCTTGGGCGCAGAG	1467
Db	901	GGGGGCGGCGCCGTGATGCTGATTTGGGCAAGATCGCGCTTCAACGCTTGGGCGCAGAG	960
QY	1468	CTGGGTTTTGCGGCTGTGGCATTTCCGTGGCGGCGGAGCTTAAACCTTGGCATGTGTCC	1527
Db	961	CTGGGTTTTGCGGCTGTGGCATTTCCGTGGCGGCGGAGCTTAAACCTTGGCATGTGTCC	1020

Qy	1528	ATCATGCACACCTTCGCGCCGCGCAATTTGTGCGACGCGCTGAGAGCTGGCATCCGCAAGC	1587
Db	1021	ATCATGCACACCTTCGCGCCGCGCAATTTGTGCGACGCGCTGAGAGCTGGCATCCGCAAGC	1080
Qy	1588	TTCGATTGCGGCGACCTTGTCCAGAGCTTATGACGCTCCCTCCGCCACGCTTTCGGTCAAT	1647
Db	1081	TTCGATTGCGGCGACCTTGTCCAGAGCTTATGACGCTCCCTCCGCCACGCTTTCGGTCAAT	1140
Qy	1648	ATTGGTCAGGGCGGTGGCGGTGGCGCGCTGGCCATCTGCGCCGCGATCTGTCTACGCG	1707
Db	1141	ATTGGTCAGGGCGGTGGCGGTGGCGCGCTGGCCATCTGCGCCGCGATCTGTCTACGCG	1200
Qy	1708	GCCGAAAAACCGTGGCTGTCTCCGATTGCGACCGAGGGGGGCTTCGCGCATCTCTTCGCG	1767
Db	1201	GCCGAAAAACCGTGGCTGTCTCCGATTGCGACCGAGGGGGGCTTCGCGCATCTCTTCGCG	1260
Qy	1768	GACACCAACACGCGCGGGAATCATATGAGCGACCAAGCGGTGACGCGACGACATTTTA	1827
Db	1261	GACACCAACACGCGCGGGAATCATATGAGCGACCAAGCGGTGACGCGACGACATTTTA	1320
Qy	1828	AGCCAGAGGCTTATCGACGGGATCTGCGCCGAAACCGAGACATTTTGTGAGAAAAATTCTC	1887
Db	1321	AGCCAGAGGCTTATCGACGGGATCTGCGCCGAAACCGAGACATTTTGTGAGAAAAATTCTC	1380
Qy	1888	GGCAATTCAGCAACGCCCTCTCTCCGAATTTGGATTAACATTCGAGAGCGCGGAGCGCGAC	1947
Db	1381	GGCAATTCAGCAACGCCCTCTCTCCGAATTTGGATTAACATTCGAGAGCGCGGAGCGCGAC	1440
Qy	1948	AGTGCCTTCAAGATTGAGCGCTTACGCGCAG	1980
Db	1441	AGTGCCTTCAAGATTGAGCGCTTACGCGCAG	1473
RESULT 11			
AP005217/c		300750 bp	DNA
LOCUS			linear
DEFINITION		BCT 20-DEC-2002	
ACCESSION	AP005217		Corynebacterium efficiens YS-314 DNA, complete genome, section 4/11.
VERSION	AP005217	BA000035	
KEYWORDS	AP005217.1	GI:23492722	
SOURCE			Corynebacterium efficiens YS-314
ORGANISM			Corynebacterium efficiens YS-314
REFERENCE			Bacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Corynebacterium.
AUTHORS			1 Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y., Ikeo, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y., Ueda, Y., and Sugimoto, S.
TITLE			The entire genomic sequence of Corynebacterium efficiens YS-314
JOURNAL			Published Only in Database (2002)
REFERENCE			2 (bases 1 to 300750)
AUTHORS			Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., and Director-General of Biotechnology Center.
TITLE			Director-General of Biotechnology Center.
JOURNAL			Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan
AUTHORS			(E-mail: bioemite.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424) Kawarabayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan
COMMENT			Nakamura, Y., Ikeo, K., Suzuki, M., and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan
			Itch, T. is at the Japan Biotechnological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan
			Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan
			Nishio, Y., Ueda, Y., and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan
			The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.
			Location/Qualifiers
FEATURES			
source			1. 300750
			/organism="Corynebacterium efficiens YS-314"
			/mol_type="Genomic DNA"
			/strain="YS-314"
			/db_xref="taxon:196164"
			252..620
			/note="CE0885"
			/codon_start=1
			/transl_table=11
			/product="hypothetical protein"
			/protein_id="BAC17695.1"
			/db_xref="GI:23492723"
			/translation="MDQAPSCIVDCVHYIFXDTPVPMILFPVGTSSGNRKIVLDFNHYFLANKPRILITPLMSGCCLFQWVIGLFTPMHTHYVCVGFGEFDPSSGRLAACLTHASRTMKPCDQGLVANG"
			529..1936
			/gene="rna_16S"
			529..1936
			/gene="rna_16S"
			/product="16S ribosomal RNA"
			/note="CERNA01"
			complement(2042..2215)
			/note="CE0886"
			/codon_start=1
			/transl_table=11
			/product="hypothetical protein"
			/protein_id="BAC17696.1"
			/db_xref="GI:23492724"
			/translation="MSQTPNSMPPTYHHGLGDVAFADPWWIHPITTCWQPHPGSNTLNPPHSHSPWVY"
			2399..5481
			/gene="rna_23S"
			2399..5481
			/gene="rna_23S"
			/product="23S ribosomal RNA"
			/note="CERNA02"
			complement(3808..4302)
			/note="CE0887"
			/codon_start=1
			/transl_table=11
			/product="hypothetical protein"
			/protein_id="BAC17697.1"
			/db_xref="GI:23492725"
			/translation="MQAPLPLRGRFAEFLNHSSEPERLSITLYITTCVGLGCGNNHIAARFSROYRITETPRLATHASNTMRGFTNTEPHLTTPQSTWMLPHCVTSLSLYYQIRSHARHHQPEGRHVLWVSTIDPWPARTIGYNINPLSIDVACRPLRPLRLTLGRLT"
			complement(4394..4594)
			/note="CE0888"
			/codon_start=1
			/transl_table=11
			/product="hypothetical protein"
			/protein_id="BAC17698.1"
			/db_xref="GI:23492726"
			/translation="MRQGSRYAIPAGRNLPDKEFRYLRMYITVTAYWGLNSQRLTQLTGPLNPARGKQSYVINL"
			5606..5725
			/gene="rna_5S"
			5606..5725
			/gene="rna_5S"
			/product="5S ribosomal RNA"
			/note="CERNA03"
			5925..7088
			/note="CE0889, similar to AX063829-1 CAC25

CDS
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAC17700.1"
 /db_xref="GI:23492728"
 /translation="MTFSFDVDEAVARRNNEFLDRTKRLQISAEFFGLVLLAIGVL
 YLLADGAVGMWVLLVMGIMALSFMVLPVLPVQVGSASLNTYPLVAVVAEINPR
 DEISLALVTNTDELTPRMGLAVRTINRLGVDRDLGERVSVASVSGRTTLDODHW
 DEISPMPTMGTTDADVVRARAKTIPIHDMNMLEKRNRLDEVKTRMLLVL"
 complement(7180..7821)
 /note="CE0890"
 /codon_start=1
 /identity: 75 in 547 aa"
 percent

CDS
 /transl_table=1
 /product="putative DNA helicase"
 /protein_id="BAC17701.1"
 /db_xref="GI:23492729"
 /translation="MAFGDGPLVQSDKTVLLETGHDLAGEARAALAPAELEAREH
 VHTYRTPLALNNARAAGHDAEQVVDLERSFPVPOPLVDVAGTMSRGRVMMK
 HPAGLLSESEPALIVEISRRKIKPMVGOQDDEPTIAVHSEBGRKOKELIKGMP
 AEDLAGYDEAHPIALSTREVDMSLRDQQAASPFEGSGCVVLPQKAGATVGA
 ASMAQAQTTILVNTVAGRWKMDLARTITLDEIGESGKESIRPVATYQV
 VTRTGERKALELFDSDWGLIYDEVHLLPAPFRMTSDQSSRRRLATVATVRED
 GREGVFSLIGPKRYDAPMKDLAEGFATADCEVTRTMDTAEWMTATATLVRD
 LAATATKTRVVRILIEHQGQPTLIIGAYDQLEBELGEPALVLDGTSKKKGL
 FDRPSELNVLRVSVKANSIDLEPAVAIVSGTFSRROEARLRLRLPKADGG
 EAHFYSVSRDLTEVAHRRPFLAEGVAAIRIIDADDLEPTTEKE"
 complement(9548..11851)
 /note="CE0892, similar to AB006976-11|AAK45126.1| percent
 identity: 36 in 741 aa"
 percent

CDS
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="BAC17702.1"
 /db_xref="GI:23492730"
 /translation="MPCFPFQIRQTKVTSAPKGLDMMKDSVPVTLTGMLEQDLD
 QLSGILNRPTALPLPGLSGIARLQLRASIQAVKLSLAEVLAEAAANAGAL
 NPITAPVEVERLHRLRHSAAVPAEIOGEALNRLROLALVGERPFIILEPTMAEL
 TNMOLLPEHGRSLTRGEASOSIEALPARHRIITVLTNSGILGTRPAALSADARP
 IPQILKAGLSRVDQYKLPQVYRNVIEGQEDTMRVVPVEPSAVPGADGILAGL
 EVARLQRMILDALSTAPASTLKEGALGVMTVRLTRELQVDEQVAVASLGSGL
 RRGVDDPLPADDDGDYIAPVADENMLQATLAEQALMGMTQTVAAMLVSEADE
 NKPTIHLSPASRRDALPETRAMILOSILTRVEEDLEADLPFHHTLAASRIPERITR
 LLEBARWIGAVAGGVTSPARVLTQAPAGVPEIADIAVAPAPVNDVYFIQADYV
 MYPGLOPEMOKITIGQLADLSPGLASTYRSEASLRHAMDGLTATEIDPLTHSA
 TGLPQSVAYLLQDILARRHGLRGGPALCYTILSDSDALLHSAVEAAGALGLOIATAV
 AVASLLEVTALRAQMGQPAEDALGASLDLAPAPAPVPAEOPRSAGPDESKIRA
 AVAALIREEVARTGTVDQPTLAVLQAAVRGQRTVTLGFDVKQGVAVHRTVPLTVNA
 GQVADLDESTGAVHFMRLRITVEIVD"
 complement(11836..12036)
 /note="CE0893, similar to AL583924-91|CAC31110.1| percent
 identity: 50 in 62 aa"
 percent

CDS
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="BAC17703.1"
 /db_xref="GI:23492731"
 /translation="WGNMANVEKGYNDPAMPDHDPSGHVVTETLIAPAGASBPMD
 DMEFPVPAETIGYVHYTRINR"
 complement(12229..12822)
 /note="CE0894, similar to AX133781-1|CAC38957.1| percent
 identity: 71 in 195 aa"
 percent

CDS
 /transl_table=1
 /product="putative cold shock protein"
 /protein_id="BAC17705.1"
 /db_xref="GI:23492733"
 /translation="MPVETKWDPERGFEVSNPGBDFGVKOVLPKGVTEHOGQ
 RIEPFAAGRGAPOLRVKVELPRRRRQHHTYKPEBLNGIISDWTMBEGTVPALIK
 GHYPRKAGEVQAKILRAVAKELEA"
 complement(13881..14408)
 /note="CE0895, similar to AB006977-4|AAK45135.1| percent
 identity: 52 in 127 aa"
 percent

Query Match 32.6%; Score 691.4; DB 1; Length 300750;
 Best Local Similarity 65.4%; Pred. No. 146;
 Matches 1046; Conservative 0; Mismatches 546; Indels 7; Gaps 2;

QY	528	TATGTTGGGGGATGGAACACACTTACGATTCAGCTCATAGACTCGGTTTGGACCC	587
DB	26058	TATGTTGAGACATGACACGAACTCCGCGGATCTGATCGATTCATTTGGACAA	25999
QY	588	TGACGCTTCATTTCTTGAATGAACCTCCCAATATGACACCTCATCAAGCTATGC	647
DB	25998	GACACGTTTCGCTTCGAGACACTCCCGAGTACGGGAGATGATGATGATGATCCG	25939
QY	648	AGAGACTTGGAGCGGGCTCCAGAGCAAGCCAAATGCGATGAATGATTAATCTGAGA	707
DB	25938	GGAGCGCTTCCCGCGCGCGCGGAGAAATCCGGGTGATGATGAGCGGTGATCAGGGTGA	25879
QY	708	AGGACCGTGAAGGCAATTCGGTACCGGTAATTTTCCGATTTTCTTCTCGGCGG	767
DB	25878	GGGAACTGAGAGGCGTCCCGGTGCGTTCATCGTCCGAGTTCCGCTTCTCGGTGG	25819
QY	768	TTCTTTGGGACGCGTGGGTGGTGGCATGATGAAGCCATTCACCGCGCACAGACT	827
DB	25818	TTGATCGGTGCGGACCTCCGCTGCTCATACAGGCTCATCCCGGCGCACCGGGA	25759
QY	828	GAACCTCCACTGCTGCTCTCCCTGCTCCGCTGCTCCGCGATGACAGAAACATCG	887
DB	25758	ACGCTCCCGCTGTTGATCTCGCGTGTCCGGTGGACACCGGATGACAGAGGACATCC	25639
QY	888	AGCTTTGTCATGATGATGTCATTAACCGCGGCTGTGACGCTCACCGGAGCGCATTT	947
DB	25698	GACGTTCCGATGATGATCTCATACACCGCGCTTACCGGACAGAGACGCCACCT	25639
QY	948	GCCGTTCTGATGATTTTGGCAATCCAGATGGGTGGCCCATGCGCTTGTGGGGTTC	1007
DB	25638	GCCGTTCTGATGATCTCATACCGCATGAGGTGGGTGATGCGCTTGTGGGGTTC	25579
QY	1008	ATCTGGGACATCTCATTTTGGGAACCGCGGCGGAGATAGATTCCGGGTCCCGGT	1067
DB	25578	CGCCGCGCATCTTACCTATGCGGAGCGGATGACCTGTGGGTTTCTCGGTTCCCGGT	25519
QY	1068	GCTGAGTTAACCATGCGGATGCGCTTCCAGACGTTGTGACAGAGCGGAGAAATTTGT	1127
DB	25518	GCTGAGTTGACACCGGACCCCATGCGGAGGCGGTGACGCGGCGGAGAACCTGCG	25459
QY	1128	GAACAACTGATGATTTGATGAAATTTGTCCGCAATTTGGTGTGACGCGGTGGCAA	1187
DB	25458	CGGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	25399
QY	1188	AACCTCAAGGTTATTCAGCGGTAGAGGCAACGATGCTTTTCTCCAAACAATCCCTGG	1247
DB	25398	GATCATGAGGTTGTTGGGAGGCGGTGCGCAGGACGCGCCCTGATCTGCGCGACCTGG	25339
QY	1248	CGTGGCACTTCCGATGATGAGAGGATGCGGCTTCTGATGATGATGATGATGATGAT	1307
DB	25338	GAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	25279

```

Qy 1308 CGGGAGATTAATGAAACGTTGGGGGACAGCTGCTAAGCTTTTGTGGCGGTGCTGG 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25278 GGGCGATCTGGTGTGCTGCTGTGGGACGAACCTGATGACCTGTCCGGCAGAGGTGAGG 25219
Qy 1368 CGCATTTGAGCCCGGCTGTGGCGTTGCCCTGGCGGCATCGGGGGCCGGCCGTGTGCT 1427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25218 TCGCATCTCGGGGCGACCAAGGTGGCTGGCCCGCATCGGGGGCCGTCCGGTGTGCT 25159
Qy 1428 GATTGGGGAGAGA---TCCCGCTTACCGTTGGGGCGCAGAGAGTGGCTTTTGGCCGTG 1484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25158 CATCGGAGAGACCGTCAACACTCCCGTGGGCGCGCCGACATCGCTTTCGCCGCGCG 25099
Qy 1485 TGGCATTTTGTGAGCGCGCGAGCTAAACCTGCCATCTGTCTCATCATCGACACTCCGG 1544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25098 CGGTATCCGCTCTGGCGCAGGAGCTGACGTGCCATCATCAGCATCATCGACCCCGGG 25039
Qy 1545 CGCCGAATTTGTGACGAGCGGCTGAGAGCTCGGCATCGCAAGCTGATTCGCCGACCTT 1604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25038 CGCGAGCTGTGCGGGAGGCGGAGAGAAAGGGCTCCCGGTTCCATCGCCGACCTTT 24979
Qy 1605 GTCCAAAGTTATCGACGCTCCCTCCCAACGCTTTCGTCATATTGTGCAAGGGGTGG 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24978 GGGGAGCTGTGCGACGCTCCGGTCCCAAGCTTCGTCATCTTGCGCAGGGGTGTGG 24919
Qy 1665 CGGTGGCGCGCTGAGCATGCTGCGCGCATCTGTCTACGCGGCGGAAAGCGCTGAGCT 1724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24918 AGGCGGTGCGCTGCGCATGCTGCTCCCGCAGCCGGGTGTGGGGCGAGAAAGCTGCT 24859
Qy 1725 GTCCGATTTGCCACAGAGGGGCGCTCGGCATCTCTTCGCGCAGACCAACGAGCGCG 1784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24858 GTCCCGCTGCGCCCGGAGGGGCGCTCGGCATCTACCGGTGACACCTCCACAGCGCC 24799
Qy 1785 GGAATCATAGAGCAGCAAGGCGTGCAGGCGCAGCAGCATTTTAACCAAGGCTTATGA 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24798 GAGATGATGAGAACCGCAGGGGGGTCCGCCATCGATGCTGGCTCCGGGATATGA 24739
Qy 1845 CGGATCTGTGCGCAACCGACACTTTTGTGAAGAAATTCGCGCAGCAATCAGCAAGC 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24738 CGGATCTATCGCGAAGCAGATTTTCATCGGGAGGTTGTGGCGGAATTCAGCAAC 24679
Qy 1905 CCTCTCCGAATTTGATTAACATCCGAGAGGGCGGAGCGCAGTCCGCTTACAGATT 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24678 CCTCACAATTTGGAACCAATCCGAGAGGGGTGTGGGAAAGCCCTTCAGGACATA 24619
Qy 1965 TGAAGCTTAGCGCAGTAAGAAATTAATGCGCTGATCAATGATGATGAACAACAAGG 2024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24618 TGAAGCGCTGGC---TAACTGAACTAAGGCGCTGATCAGTGAATGAAGACACAGG 24563
Qy 2025 TACGCGCAGACAGTGGTGGCGGAGAACCTCAGGCGCTTAAGCAGCTCTGGCGGAATGG 2084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24562 TGGCGCGGAGACGTGGTGTCCATACCTTCGCGGCGGTAGGCAAGCTTCGCGGGGATGG 24503
Qy 2085 TCACTGACGAGCTCGCGACCTTCATGCTCTGGAATTC 2123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24502 TCAGCTGACGAGCTCGCCCAACTTCATGCTCCGGGATGC 24464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
CGG/TC C 3013 bp DNA linear BCT 17-FEB-1995
LOCUS C:glutamicum glt gene for citrate synthase and ORF.
DEFINITION X66112
ACCESSION X66112
VERSION X66112.1 GI:505580
KEYWORDS citrate synthase; glt gene.
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 (bases 1 to 3013)
AUTHORS Eikmanns,B.J., Thum-Schmitz,N., Eggeling,L., Luedke,K.U. and
Salm,H.
TITLE Multicentric sequence, expression and transcriptional analysis of the
Corynebacterium glutamicum gltA gene encoding citrate synthase

```

```

JOURNAL Microbiol. 140, 1817-1828 (1994)
REFERENCE 2 (bases 1 to 3013)
AUTHORS Eikmanns,B.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1992) B. Eikmanns, Institut f. Biotechnologie 1,
Forschungszentrum Juelich GmbH, I B T, Postfach 1913, 5170 Juelich,
FRG

FEATURES
source location/Qualifiers
1..3013
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/strain="ATCC13032"
/db_xref="taxon:1718"
717..>2151
prim_transcript
838..2151
/gene="glt"
838..2151
/gene="glt"
CDS
/EC_number="4.1.3.7"
/codon_start=1
/transl_table=11
/product="citrate (si)-synthase"
/protein_id="CAA46902.1"
/db_xref="GI:505581"
/transl_table=11
/transl_id="CAA46903.1"
/db_xref="GI:505548"
/db_xref="SWISS-PROT:P42458"
/translation="MEKPIELPVPAPEDLVISDII VEGBARPEGEVEVHYGVDF
FETGESEDSMDGQTSOPPLNGLIAGMDEGIPGMKVGRRULTIPEAAVYGESEGH
PLSGRTVFIIDLSA"
2276..2632
/feature="ORF"
putative binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAA46903.1"
/db_xref="GI:505548"
/db_xref="SWISS-PROT:P42458"
/translation="MEKPIELPVPAPEDLVISDII VEGBARPEGEVEVHYGVDF
FETGESEDSMDGQTSOPPLNGLIAGMDEGIPGMKVGRRULTIPEAAVYGESEGH
PLSGRTVFIIDLSA"

BASE COUNT 716 a 818 c 799 g 680 t
ORIGIN
Query Match 24.3%; Score 515; DB 1; Length 3013;
Best Local Similarity 100.0%; Pred. No. 2.3e-106;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1609 AAGCTTATCGACGCTCCCTCCCAACCGTTTGGGTATTTATGTCAGGGCGCTGGGGGT 1668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3013 AAGCTTATCGACGCTCCCTCCCAACCGTTTGGGTATTTATGTCAGGGCGCTGGGGGT 2954
Qy 1669 GGGCGCTGGCGCATGCTGCCCGCGCGATCTGTCTACGCGGCGGAAACGGTGGCTTCC 1728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2953 GGGCGCTGGCGCATGCTGCCCGCGCGATCTGTCTACGCGGCGGAAACGGTGGCTTCC 2894
Qy 1729 GCATTTGCACAGAGGGCGGCTCGGCATCTCTTCGCGCAGCAACCAACGCGCGGAA 1788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2893 GCATTTGCACAGAGGGCGGCTCGGCATCTCTTCGCGCAGCAACCAACGCGCGGAA 2834
Qy 1789 ATCATAGAGCAGCAAGGCGTGCAGGCGCAGCATTTTAAGCCAAAGGGCTTATCGAGGG 1848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2833 ATCATAGAGCAGCAAGGCGTGCAGGCGCAGCATTTTAAGCCAAAGGGCTTATCGAGGG 2774
Qy 1849 ATGTCGCGGAAACCGAGCATTTTGTGAAGAAATTCGCGCAATCAGCAACGCGCTC 1908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2773 ATGTCGCGGAAACCGAGCATTTTGTGAAGAAATTCGCGCAATCAGCAACGCGCTC 2714
Qy 1909 TCCGAATTGATTAACATCCGAGAGGGCGGAGCGCAGCATGCTTCAACGATTTGAG 1968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	2713	TCCGAATTGATTAACAATTCGGAGAGGGCGGGACGCAGTCCGTTCACAGCAATTTAG	2655
Qy	1969	CGTTTAGCGCAGCTAAAGAATAATTATGCGCTGATCAAATCGATGATGAACACCAGGGTAG	2028
Db	2653	CGTTTAGCGCAGTAAAGAAAATTATGCGCTGATCAAATCGATGATGAACACCAGGGTAG	2594
Qy	2029	GCCAGACAGTGGGGTGGCCGGAAACCCCTCAGAGGCCGTAACACACTCTTGCCGGAAATGTGAG	2088
Db	2533	GCAGACAGTGGGGTGGCCGGAAACCCCTCAGAGGCCGTAACACACTCTTGCCGGAAATGTGAG	2534
Qy	2089	CTGACGACGTCCGCCGACCTTCATGCTCGAATTC	2123
Db	2533	CTGACGACGTCCGCCGACCTTCATGCTCGAATTC	2499
RESULT 13 AE006979/c			
LOCUS	AE006979/c	17910 bp	DNA linear BCT 27-APR-2001
DEFINITION	Mycobacterium tuberculosis CDC1551,	section 65 of 280 of the	
VERSION	complete genome.		
KEYWORDS	AE006979 AE000516		
SOURCE	AE006979.1 GI:13880475		
ORGANISM	Mycobacterium tuberculosis CDC1551		
REFERENCE	Myobacterium tuberculosis CDC1551		
AUTHORS	Bacteria; Actinobacteria; Actinomycetaceae; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium tuberculosis complex.		
	1 (bases 1 to 17910)		
	Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,C., Peterson,J.F., Deboy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.		
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 17910)		
AUTHORS	Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,C., Peterson,J.F., Deboy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.		
	Direct Submission		
	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1..17910		
	/organism="Mycobacterium tuberculosis CDC1551"		
	/mol_type="genomic DNA"		
	/strain="CDC1551"		
	/db_xref="taxon:83331"		
	/note="clinical strain"		
	145..1326		
	/gene="MT0918"		
	145..1326		
	/gene="MT0918"		
	/note="identified by Glimmer2; putative"		
	/transl_table=1		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="AAK45164.1"		
	/db_xref="GI:13880476"		
	/translation="MPSPATVVOEFSDSYPFGHNGFPIIMPKIVSVSHSRRLUTSFVGRKLNDVRRLSDSKRLVTLTGPDSFGSGKRLLAQIAGAIARHPYGRMPDCLAYTDRCDCISIMNALGCPYOPGSAIDITLVGYINDARVLVDHCHGLDACAIIIDSILRSCEPLUILLTSTEAIEIAAARTLTPESVEILAGINRFRLVAGAAGNAVBEYTLCAATVVSHTALSGPERALLHLRAVFAFGFLDGAOVGANDEDFEGYTIGRAELVDAFVVVENNRGARGRYLLYSRVQALEKSESSEADAVLARVKHLKPQNQVVRASGGVRY"		
gene	1358..2920		
	/gene="MT0919"		

1358. .2920
/gene="WT0919"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45165.1"
/db_xref="GI:13880477"
/translation="MTAPPIAVERNRTSKVRQOQEAADVVALGKRPGLLCVPEEPRAND
LPMADADLFLMAETPTAPPLIVGALVALVSGDNGTGYLKRVSAAVAROOVAPWRR
BPHSLTISGQMSRTEVEDVDYVLSALPRAGTAEILMALYSLELHAMDLNPSPL
MOVDLIELPBGRCAYVYKYNHNLADSVYKRLQRIYVTDPRHQRPPTLMEYPAQS
VAKHTAPGSSRPDLTAGVLDQAKGVGVNVRVADDTMPRAQCRSGPDLTAHPHPL
NEPIAGASVAGSGFPPIERLNOVAEHADATINDVVALMCGGALRAYLISGALGPAL
IAMPVLSARDVAVDVFQGGPNKIGTLMGSLTHLASPYERLSAIVASRDKKAIA
GSRNOALMAGALAAPLALAMALGAVPAIPPNVTISVPGQGLYNNGARLDL
YLASPVGALALNITCGSTNQITFGLTGCRRAVPALSLITDLAHELLVGSFENG
PGRRLRACGR"

3093. .4388
/gene="WT0920"
3093. .4388
/gene="WT0920"
/note="similar to PID:2440108; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="citrate synthase"
/protein_id="AAK45166.1"
/db_xref="GI:13880478"
/translation="MADTDPTATLRYGGSIIIDQIYHATEGADGIALGPLANTGHTT
FDVGVNAAAKSSITIIDDGILRRGPIQLAKSFTFIVCYLLTYGELPDDQ
LAQFTGRJORTTMAHEDLKRFDFDPNAPHPMVLSSVVALSAVODALDPMDNQV
ELSTRIALAKPTIAAAYAKKSGVQPELYDNSLTIVENTRLTFGPBAPYQADQV
VRALDMLFIHADHONCSTSVLRVSSRANFTISGGTNALMGPIHGNQADYLE
MLEGIRPSGDDVSEFYKRYKRRBAGVYLMKFGKRYKYNDRPRRIYKEDQDKLALG
GDSLILGAKLEBALITDDFTIERKUYPNVDYTGILTYALGEPPTMFVLPALGRL
PGMLAHMKEHDEDSLIGRPDIYGTGTRSDYVTIDAR"
complement (4429. .6036)
/gene="WT0921"
complement (4429. .6036)
/gene="WT0921"
/note="similar to PID:1001734 PID:1001724; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, putative"
/protein_id="AAK45167.1"
/db_xref="GI:13880479"
/translation="TMSDHRDPDVVVVGGNGI,VAAAYILRAGLRVRLLERLIAQCG
AAVSIQAFDGEVALISRYVLSILPERIYADAGAPRLARFPSSYTPAPAPASGG
LLIIPTEPPRAHIALGAAPDAHGFAPRRCVLTARLMPLELITLREBQARDI
VEVGGHAAAMQAMDEPRTIGHALAGAVANDLIRGYITADLIGTPARKNHEPSLMONT
CFYHLVGGGTGVNHPITGMSYSTSLATPAARHGAELVTGADVFDALPDGVRVHS
DGSAGAHLVGRGRLVLTGPVAVLASLIGPVALLAPGAQVKVNVVRLPRLRDVS
TPQQLPFGTGHVETVETQDLDAASQASGRLPDLPCBAVCHSLTDPSTLSARLDSG
AQQITVGGTLPHPVFGDTGELARLTAAYLASINVLAEPIQDVTLPATQSPCTET
TTTLDLRTTGTCGNIFHGLASMPFANDNDPLDPFARQGVATDHERIMLCSGAR
GGANVSGCHNMAAVYLAASRRKSP"
complement (6062. .6325)
/gene="WT0921.1"
complement (6062. .6325)
/gene="WT0921.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45168.1"
/db_xref="GI:13880480"
/translation="MGKGRKPTSEHTLAHRLVAEEKALRAQLRHHGSISSBEOQL
RRIEHLDDQWDLRQRRLRQTGSDPREAVVRPADQVGYTG"
6433. .7413
/gene="WT0922"
6433. .7413

Query Match

```
/gene="MT0922"
/note="identified by match to protein family HM"
/codon_start=1
/transl_table=11
/product="OmpA family protein"
/protein_id="AAK45169.1"
/db_xref="GI:13880481"
/translation="MASHKAGIGOTPATTDARTOKYRSGRPMILGAVVPIILIA
IGVAFERPOSVGPTGVLPILPTSTRASALSLISISGTVTILIGDPDEBA
KKAATMALGILAPGVNVDQIHDPVVASLSPSSAEVFTASVPIPDGLKVEDTV
TILGTAPSESHKDAVRAATSTPMKIVNIEVYQAPGPPASGACADLQSAINAV
TSGPIAFVNDGASLIADYEILNRVADKLKACPDARTINGTYDNTGSGSIVIPLSAQ
RAKIVADYLVARGVADHIAIVLGSVNPISNATPEGRAKNRVEIYV"
/length=7578
/gene="MT0923"
/length=7426
/length=7578
/gene="MT0923"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45170.1"
/db_xref="GI:13880482"
/translation="MDFVIQWSCYLLAFLGSAVAVVVTLSIKRASHDEGAAPSA
AERTAQ"
/length=8105
/gene="MT0924"
/length=7575
/length=8105
/gene="MT0924"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45171.1"
/db_xref="GI:13880483"
/translation="MMEHVHWLAGIAFTLGNVLTSLMVEPVEHQVLKXSGSSA
KSKPTARKPAVSGTKESEPTAKTKATESAEOIIVAGEPAEPIPVAGEPARI
PVVPAYPGPGSARAGADSGPGQVLVGRSDTRLVYTPEDPTDPTAQVWPGDEBS
AARAFPTMRKSTRRT"
/length=8122
/length=9531
/gene="MT0925"
/length=9531
/length=9531
/gene="MT0925"
/note="similar to PID:2440113; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAK45172.1"
/db_xref="GI:13880484"
/translation="MAGCCIPSAESDSCVCSGDAQCNILRIAPRPSLRTPVVA
TAIGAIPVLIVGVVWVGINDRKERDRLDEAGAIIPVVRGIDEIRSNDDO
ALITVRGNVTKNSDITLPLKDDYATTVYRVYRVEIIEGPEITSVAVATYD
ATVATNLNHRVLICTFAIGAAVFMILAAFAVREFQLAQTSIDAGDAPRV
EVHGASEAIEIAEMRGWLQRIWNEONTKEALASARFAVSSHEIRPLTARTNL
EVLSTLDPDQKEVLNDVIRTOSRIATLSALERLAQGLSTDHVPDIDTLDL
RAADPAARIYEDLVSLVSPICIIIVGLPAGRLAVDNAIANVKGAGATITVLSA
SRAGEIAIDNDSGVPEGRQVVFERSRSTASHSGSLGLAVVAQQLHSGTAS
LENSPLGARLYVRLRPS"
/length=9473
/length=10183
/gene="MT0926"
/length=9473
/length=10183
/gene="MT0926"
/note="similar to GP:2440114; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-binding response regulator"
/protein_id="AAK45173.1"
/db_xref="GI:13880485"
/translation="MGGMIDGVTSFPRVLVVDSDVLAISERGLSLGFEVAVAVDGA
```

14.7%; Score 312.6; DB 1; Length 17910;

```
Best Local Similarity 54.0%; Pred. No. 3e-60;
Matches 717; Conservative 0; Mismatches 569; Indels 21; Gaps 3;

QY 572 ACTGGTTTGGACCTTGACAGCTTCATTCTTGGAAATGAAACCTCCCAATATGACAACC 631
Db 11770 ACGGGGTCTAGACCGGGGATCTTCTCGACGTGGATATGACAGCCCTGGCGGTGCCG 11711
QY 632 TCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAACCAAGGCCAAATGCGATGAT 691
Db 11710 TAGCCGACTCTTAAGCCCGGAGCTGGCCCGCTCGGGCGGCCACCGGGCGGACAAAT 11651
QY 692 CGGTAATTAATCGAGAAAGGACCGTGAAGGCAATCCGGTAAGCCGTATTTGGCGAAT 751
Db 11650 TCGACTTCTGGGGGGGTTCGATTTGGGGTGGGAGGCGGACGATACCGCGCCGTG 11531
QY 812 ACCGCGCCACAGACCTGAAACTCCACTGCTGCTTCCCTGCTTCGGTGGTGGCGCA 871
Db 11530 AGCGGGGACCGCGACCGGCTGCGCTACCTGGCGTCAACCAAGCTGGAGGACCGCA 11471
QY 872 TGCAGAAACAATCGAGCTTTGTATGATAGTGGTCCATACCGGGCTGTGACAGCTG 931
Db 11470 TGCAGAAAGCACGCTGCGCTTCTGCAATGGTGAAGATCGCTGCGGCATCAGCTGC 11411
QY 932 ACCGAGAGGCGCATTTGCGCTTCTGCTGATTTATTTGGCAATCCACAGATGGTGGCGCA 991
Db 11410 ACAACGAGGCGCGCTGCTTCTGCTGATTTATTTGGCCATCCGACAGGGTGGAGTTT 11351
QY 992 TGGCTCTGGGGTTATCTGGGCACTCTCTTTGGGAAACCGGGCGCGAGATAGTT 1051
Db 11350 TCGGTGTGGGGCTCGCTGGGGCATCTCACCTGCCGAGCGGGCGCTGATGGCT 11291
QY 1052 TCGTGGTCTCGCGTGGTGGAGTTAACACCTGGGATGGCTTCCAGACGGGTGACAG 1111
Db 11290 TTTGGGACCAAGGCTATGAGTTGCTATGAGCAACCTTCCATCGGAGTCCAAA 11231
QY 1112 AGCGGAGAAATTTGGTGAATACTGGTGTGATTTGATGAAATTTGTGCGCATCTCAATTC 1171
Db 11230 CCGCGAGAAATCTACCGCGGATGGATATCAGCGGCGTGTGACATGACCGGCTAC 11171
QY 1172 GTGAGCGGTGGCAAAACCTCAAGTTATACCGGTGAGGCAAGGATCTTTT 1231
Db 11170 GACGATGCTGATCGTGGTTCAGGCTCATACACCTCCGAAACCGCTTCCGGCAC 11111
QY 1232 CTCCAACAATCTCTGCGTGGCA---CTTCCGGTATGAGAGGCGATTTGCGCTTCTG 1288
Db 11110 CGGAGACGCGCGCGCGCTGACCGATGTCACCGTGGGACCTGGTGGCATCGCGCC 11051
QY 1289 ACCGCGAGAGCTGGAATCGGGAGATTAAGAAACGTTGGGGGAGACGCTGTCAAC 1348
Db 11050 GCGCGGACCGCGCGGCTGACGACCTACTGGAACACGGCGCACGACCGGGTGTGT 10991
QY 1349 TTTCTGTGGCGTGTGGCGCATTTAGCCCGGCTGTGGCTTGGCCCTGGCGCGCATG 1408
Db 10990 TGTAGGAACGATCAAGGGAAGCG---GCAACACGCTGTGGCTGGCGCGCGCTTTG 10934
QY 1409 GGGGCGGCGCGGTGCTGATTTGGGCGAG-----GATCGCGCTTCAACGC 1453
Db 10933 GCGGCGCAACCGAGGTGCTCTCGGCAAGAAAGGCGAGTACGCGCGGGGAGACACTG 10874
QY 1454 TTGGGCGGAGAGCTGCTTTTGGCGGTGCGCATTTGCTGGCGCGCGAGCTAAAC 1513
Db 10873 TCGGGCCCGCTGTGCTTACCGGAAGCCGACGCGGATGCGCTCGCGCGCGAGCTGTGCC 10814
QY 1514 TCGCGATCTGTTCATCATGACACTCGCGCGCGGCAATTTGTGCGAGGCGCTGAGAGC 1573
Db 10813 TGCCTGTGTGTGCTGATTCAGCGCGCGGACCGCGCTGTGTGCGCGGACGCAAG 10754
QY 1574 TCGGCAATCCCAAGCTGATTTGGCGGACCTTGTCAAGCTTATGACAGCTTCCCTCCCA 1633
```


Db	10753	GGGGGCTGGCGGACAGATCGCGCATTTGCCATGGCCGAGCTGTCAGCGCTGGATTAACCCCGA	1069
QY	1634	CCGTTTGGGATTAATTGTGTAGGGCGTTTGGCGGTGGCGGCTGGCCATCTGCGCCGCG	1693
Db	10693	CCGTGTCGATCCTGTGTGGGCCAGGGGACAGCGGCGGGCGGCGCTGGCGATTTTCCGCGCG	1063
QY	1694	ATTCGGTTACGCGGGCCGAAAAACCGTGGCGTGTCCGATTTGCCACACGAGGGGCGCTCGG	1753
Db	10633	ACCGGGTGTGGCGGCACTCCAGCGCTGGCGTGGCGCCCTTGTCCCTCCGAAGAGACGAGCG	10574
QY	1754	CCATCTCTTTCGCGGACACCAACGACGCGCGGAAATCATAGAGCGCAAGGCGTGCAGG	1813
Db	10573	CGATCGTGTTCGAGAGACATGCTCTATGCGCGGCAACTGCTGCGCGCCCAAGCATCCGAT	1051
QY	1814	CGCAGCGCATTTTAAGCCAGAGGCGTTATCGACGGGATGTGCGCGAAACCGAGCACTTGG	1873
Db	10513	CGGCGGACCTTACGTAAGTCGGGGATGTGTCGACATCTGTGCGGAGTACCCGACGCGG	1045
QY	1874	TTGAAGA 1880	
Db	10453	CAGACGA 10447	
RESULT 14			
LOCUS	MTX313	37630 bp	DNA linear BCT 03-AUG-2001
DEFINITION	Mycobacterium tuberculosis H37Rv	complete genome; segment 41/162.	
ACCESSION	Z73101	AI123456	
VERSION	Z73101.1	GI:3261565	
KEYWORDS			
SOURCE			
ORGANISM	Mycobacterium tuberculosis H37Rv		
	Mycobacterium tuberculosis H37Rv		
	Bacteria; Actinobacteria; Actinomycetales;		
	Corynebacterineae; Mycobacteriaceae; Mycobacterium		
	tuberculosis complex.		
REFERENCE	1 (bases 1 to 37630)		
AUTHORS	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gae S., Barry III C.E., Tekari F., Badcock K., Baah D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandrem M., A. Rogers J., Ruter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S. and Barrell B.G.		
	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
	Nature 393 (6685), 537-544 (1998)		
JOURNAL	98295987		
MEDLINE	9634230		
PUBMED	2 (bases 1 to 37630)		
REFERENCE	Parkhill J.		
AUTHORS	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75124 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk		
JOURNAL	On Jun 27, 1998 this sequence version replaced gi:1311009.		
COMMENT	Notes:		
	Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.		
	(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.		
	Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.		
	CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation		

```

FEATURES
Source
codon . If this cannot be identified we choose the most upstream
initiation codon.

Location/Qualifiers
1. .37630
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="y31"
complement(54. .1214)
/gene="Rv0874c"
complement(54. .1214)
/note="Rv0874c" (MTCY31.02c), len: 386. Unknown, very
similar to Mycobacterium tuberculosis MTCY20H10.9 (383 aa
E):0, 81.5% identity in 383 aa overlap) and to
hypothetical proteins from SPUB2616.1 Synecchococcus sp.
(280 aa; E): 6.3e-26; 35.2% identity in 264 aa overlap)
and SYCSLLH102 Synecchocystis sp (447 aa; E): 1.1e-18;
29.5% identity in 400aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0874c"
/protein_id="CAA97403.1"
/db_xref="GI:1314013"
/db_xref="SWISS-PROT:Q10536"
/translacion="MRIGVGCTTPDARQAAVEAAGARDELAGEAPSLAVLGSRAH
TDRADVLASVLMIDPDPALVGCIAQIVAGRHEIEBPVVWMLASGLAETPOLDD
VRSGGALLTGRFDRPTARDLHLLPDPYTPSPVLSLEHPTDLPGTAVGAVSGG
RGGRTLRFRHDVLTSGVGRVLRPGMGSPVPSGCRPIGVPIVTCAGDGLITTEIG
RPLQRLREIVEGLSPDERALVSHGLDITGVDEHDLARPGGDVYIRGLDADSTGS
IEIDEVQVAGATMQFQVRDAAGADKDLRLVERAAALPGAAQGLLFTCGRGRRMM
GVADHDASTIEELGLGPIAGFFAAGEIGPIAGRNALHGFYSNALPVDME"
complement(1314. .1802)
/gene="Rv0875c"
complement(1314. .1802)
/note="Rv0875c" (MTCY31.03c), len: 162, unknown, possibly
exported protein, highly similar toMC1B57.11 Mycobacterium
leprae cosmid B57, O33056 HYPOTHETICAL 18.1 KD PROTEIN
(162aa). FASTA scores: opt: 789 z-score: 963.4 E(): 0;
71.4%identity in 161 aa overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0875c"
/protein_id="CAA97382.1"
/db_xref="GI:1314012"
/db_xref="SWISS-PROT:Q10537"
/translacion="MKRGVATLPVLTIVLLSVAGAGAMLLVRGHGQOPEISAVSHG
HLTRVQPYLVLCNVDDDCQTPQAGELPVSERPVQLSVEVLSRAPWRLQVYQDFE
ANTSTSLFRPDTRLAVTIPIVDPDQRLTGIIVQLTLVVDSHSELNDVPAHMSVRL
IF"
complement(1799. .3445)
/gene="Rv0876c"
complement(1799. .3445)
/note="Rv0876c" (MTCY31.04c), len: 548; unknown; possible
membrane protein, highly similar to MLCB57.12
Mycobacteriumleprae cosmid B57 O33057 HYPOTHETICAL 61.7 KD
PROTEIN (579aa). FASTA scores: opt: 2850 z-score: 2863.4
E(): 0; 81.0% identity in 568 aa overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0876c"
/protein_id="CAA97383.1"
/db_xref="GI:1314013"
/db_xref="SWISS-PROT:Q10564"
/translacion="MAPTPGRRTKNSGVNHGPMQANYPDPDANYRRSRPPMPMSNRR
YSLPLEQEPERSRVPRTTRAGERITVTPRAAMSREWSRNYLLVHRAATGDAGD
KSGLTITLTPWNAFVDSAMAVLANLTLPFAASGESKSRVLLYLTITAPPAVIRA
LIGPADRLQHGRRVALLASFGKRLTAVLVLLINVDGATSPFSWVLYPCALAMVVS
KSSVSVRSVTPRPVPTIDLVKVNSSLRTVFGLLGGITACGAAAGAEVCTHLPQLE
GALFVVAVLTIGASISMRIPRWVETSGVSPVLTLSYHNRGRLRRMPEVKNLQGLPQ

```

LAOPLGRNITITSLMGCTIKKAVGFLPYLPAPVAAKHAANGVOMGLGILGAANAAG
 NFGANQTSRLQGRPAVLVRCVTLVTLVLAIAAVASLATATATITIGSSAIAAG
 ASIALDNLQDLPESNRASGFRSESITLQALWGLAGVGLVTEIHWGFIATVSLAILI
 GLAQITVSRFGDSLIPGLGGRPPVMAEBOETTRGCAVAPQ"
 complement (3457..3465)
 RBS
 /note="gagcgtgacc, potential rbs upstream of Rv0876c"
 3572..3575
 gene
 /note="gagg, potential rbs upstream of Rv0877"
 3583..4371
 /gene="Rv0877"
 3583..4371
 /gene="Rv0877"
 /note="Rv0877, (MTCY31.05), len: 262, unknown,
 equivalent to MUCB57_13 Mycobacterium leprae cosmid B57
 033058 HYPOTHEITICAL 28.2 kd PROTEIN (269 aa; E(1): 0.80.5%
 identity in 257aa overlap.)"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0877"
 /protein_id="CA97384.1"
 /db_xref="GI:1314015"
 /db_xref="SWISS-PROT:O10539"
 /translation="MTGPRESAVATVADWPEGLAALVIRGAADORAARVVFSGSEPAV
 GVLYGVSVDNQAATHRPRIARLPGYQGMOMAVYASYSAGADATITSEVLPGETALL
 AADVPMWEORVRPGLDISREDLLAPAKDPRFLVPGYTASGDQVETAEITELGRWW
 SHVWGRQASQWRHWDGYPGSGAMASTKVRCDGFFLPPLGSGIGAMFGVGNLSAD
 GAVVRYOYCGASDITPAAGSGSTPIYEPYDDGVLDITKEPAES"
 complement (4380..5711)
 /gene="PPE"
 complement (4380..5711)
 /gene="PPE"
 /note="Rv0878c, (MTCY31.06c), len: 443, unknown; similar
 to y56 MYCTU P42611 hypothetical 50.6 kd protein in hsp65
 3, region, (517 aa), PASTA scores: opt: 1044, z-score:
 954.9. E(1): 0. (47.4% identity in 397 aa overlap). Similar
 also to MTV014_3, MTC165_2, MTC198_24, MTC307_23,
 MTC148_17, MTV004_5, MTV004_3 etc."
 /codon_start=1
 /transl_table=11
 /product="PPE"
 /protein_id="CAA97385.1"
 /db_xref="GI:1314016"
 /db_xref="SWISS-PROT:O10540"
 /translation="MNFWLPPREVASRIYVAGAPAPMLAAVAVMDGIAELGMAAAS
 FELLISGLPAGSGSAMOGPAAAMAAAPLYSLWNTATATARAAGAAKAAAVYBA
 AAAATSHPLAANRNOLLSITLVNLPQONLPATATASYTEOLMAODVAMVGHGG
 ASTVASQLPWOQLLSVLPVYTAAPAAVGPALAIPLAGVENIGVGNLTGIGNTG
 KNNVSGNTGIDTNGIGNTGNANTLGNATLGNANLGNAGFNFNGNDGNTNGSNG
 ACPNLNGSGNEGSGNLGFENAGADDNTGNGSGDNTGTGFNSGDNTLTGISTVYGVAN
 SGGFNGTGHSGFPGVNSGSGFOMLGNAGSGGFGNASTSGFONAGALTRASTVA
 DSPRAPIPAPSRITVQWTRTARTARECSIRVLSHSSTAPQPKVANSQ"
 complement (5719..5721)
 /note="gga (or aagg) potential rbs upstream of Rv0878c"
 complement (5989..6264)
 /gene="Rv0879c"
 complement (5989..6264)
 /gene="Rv0879c"
 /note="Rv0879c, (MTCY31.07c), len: 91 unknown, equivalent
 to MUCB57_14 Mycobacterium leprae 033059 HYPOTHEITICAL 9.8
 kd PROTEIN (91 aa), E(1): 1.2e-25, 76.9% identity in 91 aa
 overlap."
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0879c"
 /protein_id="CA97386.1"
 /db_xref="GI:1314017"
 /db_xref="SWISS-PROT:O10541"
 /translation="MSVNSQIRBPPLPVILEVWPYIAVGALAMLVAAVAAVVG
 IASMPVYATGATGLTGLTITVWQLAAARGAGAAQAGLETYIDPK"
 6442..6873
 /gene="Rv0880"
 6442..6873
 /gene="Rv0880"
 /gene="Rv0880"

Query Match	14.7%	Score 312.6	DB 1	Length 37630
Best Local Similarity	54.0%	Pred. No. 3e-60		
Matches 717	Conservative 0	Mismatches 589	Indels 21	Gaps 3
<div> <div> /note="Rv0880, (MTC31.08), len: 143, unknown, equivalent to MLCB57_15 Mycobacterium leprae cosmid B57 (143 aa) O33060 HYPOHEPICAL.15.8 KD PROTEIN. FASTA scores: opt: 818 z-score: 980.3 E(): 0; 89.5% identity in 143 aa overlap." </div> <div>/codon_start=1</div> </div>				
QY	5712	ACTCGGTTTTGGACCCCTGACAGCTTCACTTTCTTTGTAATGAAATCCCAATATGACAAACC	631	
DB	35657	ACGGGGGCTAGACACCGGGGATCTTTGTCAGCTGGAGATAGCGAGCCGTGGCCGTGCCG	35996	
QY	632	TCATCAAGGCTATGACAGAACCTTGGAGCGGGCTCGAAGCAAGCCCAATGCGATGAAT	691	
DB	35557	TAGCCGACTCTCTATGCGCGGAGACTGGCCGCCCTGGGGGGCCACCGCGCGAGAAAT	35533	
QY	692	CGGTAAATTAATGAGAAAGGACCGTGGAGGGCAATCCGGTAGCCGTATTTTGTCCGATT	751	
DB	35537	CGGTGACAGACCGGTGAGGAGCGGTATTTGGGGGGGGGGGCGGTGGCCGTGGAGT	35478	
QY	752	TTTTCTTCTCTCGGGGTTCTTTTGGGCAACGGTCCGCTCGGTGCGCATATGAAGCGATT	811	
DB	35477	TCGACTTCTCGGGGGGCTCATTTGGGGGTGGCACGGCCGAAACGATACCGCCGCGTGG	35418	
QY	812	ACCGCGCACAGAGACTAAACTCCCACTGGTGTCTCCCTGCTTCGGGTGGCGCGCA	871	
DB	35417	AGCGGGGACCGCCGAGACGGCTGCCCTACTGGCGTCAACMACTGGGAGGCAACCCGA	35355	
QY	872	TGCAGAAAGCAATTCGAGCTTTTGTGATGATGTGTCATAACCGCGGGCTGTGCAGCGTC	931	
DB	35357	TGCAGAAAGGACCGGTGGCGTTTCTGGAATGTGTGAAGATCGTCCGCGCATTCAGCTGC	35298	
QY	932	ACCGGAGGGGCGCATTTGGCCGTTCTGTGTATTTGGCGCATCCACAGATGGGTGGCGCA	991	
DB	35297	ACAAACGAGGGCGCGCTCCCTACTGATGTATTTTGGCCATTCGACACAGGGGTGAGTTT	35238	
QY	992	TGGCTGTGGGGGTTCACTCGGGGATCTCACTTTGCGGAAACCGCGCGCGAGATGAGTT	1051	
DB	35237	TGCGGTGTGGGGTCTCGTGGGGATCTTCAACCGTCCGACGACCGGGCGCCCTGATGGCT	35178	
QY	1052	TCTGGGTCTCTCGGTGTGTGAGATTAAACAATGGGCGATGGGCTTCAGACGGTGTGCAC	1111	
DB	35177	TTCTGGGACCAACGGGTCTATGATGTGTGTATGGGCAACCCCTTCCATCCGGGTCCAA	35118	
QY	1112	AGCGCGGAATTTGGTGAATACTGGTGTGATTATGAAATTGTGTGCCACTTCAATTGC	1171	
DB	35117	CCGCCGGAATCTAACGGCGGCGCATGGGATCATTCACGCGCTGTTCACATGCAACCGCTAC	35058	
QY	1172	GTGAGGGGGTGGGAAAAACCTCAAGGTTATTTAGCCCGGTAGAGGCAAGATCGTTTTT	1231	
DB	35057	GACCGATGCTGATCTGCGTTACAGGTGTCTATTCGACGCTCCCGAACCCTTCCGGCAC	34998	
QY	1232	CTCCAAACACTCTCTGGCGTGGCA---CTTCCGGTGAATGAGGCGATTTGCGCTTCTCGT	1288	
DB	34997	CGCAGAGCCCGCGCCCTGATCCCGATGTGCGCACGTTGGGAACTGTGTGTGGATGGCGC	34938	
QY	1289	ACCGCAGAGGCTTGGAATCGGAGATTAATGAAAAGTTGGGGGCGACAGTGTCAAC	1348	
DB	34937	GGCGGACCGGCGGGGCGTACGACACTACTGCGACACGCGCGCACCGAGGTTTGT	34878	
QY	1349	TTTTTGTGTGGCGTGTGTGGCGCATTTAGACCCGGGCTGTGGCGTGTGCTTGGCGCGATG	1408	
DB	34877	TGTAGGAACCGATCAAGGCGAAGCG---GCGACACCGTGTGGGGTGTGGCCCGCTTGG	34821	
QY	1409	GGGCGCGGCGCGTGGTCTGATTGGGCGAG-----GATGCGCGTTCAACGC	1453	
DB	34820	CGGCGCAACCAACGTTGTCTTGGCGACGAAAGGCAATGAGCGCGGGGGAAGACATCG	34761	
QY	1454	TTGGGCGCGAGAGCTGCGTTTTTGCAGCTGTGGCATTTTGGTGGCGCGAGCTAAAC	1513	

Db 34760 TCGGGCCGCTGCTTACGCGAAGCCGACGCGGATGCGCTGCCGCCGAGCTGTGCC 34701
 1514 TCCGATCGTGTTCATCATGACACCTCCGGCGCGGAAATTGTGAGGGGCTGAGAGAC 1573
 Db 34700 TCCCGGTGTGCTGCTGATTCATGACCGCGGACCCCGCTGTGCGCCGACGCCAAG 34641
 1574 TCGGACATCGCAAGCTGATTTGCGCGACCTTGTCCAAGTTATGACGCTCCCTCCCA 1633
 Db 34640 GCGGGCTGCGCGGACGATCGGCAATGCTGCGCGAGCTCGTACAGCTGATACCCCA 34581
 1634 CCGTTTCGTCATTAATTGTGAGGGCGTTGGCGGTGCGCGCTGCGCATGTGCCCGCG 1693
 Db 34580 CCGTGTGATCTGCTGCGCGACGCGCGCGCGCGCGCTGCGGATGTTGCCCGCG 34521
 1694 ATCTGATACGCGCGCGAAGACGCGTGTGCTGCTGCGCATTCGACGAGGCGCGCTCG 1753
 Db 34520 ACCGGGTGTGCGCGACCTTCACGAGGTGCTGCGCGCTGCTGCTGCGGAGGACGAGCG 34461
 1754 CCATCTCTTCCGCGACACCAACGCGCGGAAATCATAGACGACGAGCGTGCAG 1813
 Db 34460 CGATGCTGTCCGAGACACTGCTCAATGCGCGCGAATCGCTGCCCGCAAGCATCCGCT 34401
 1814 CGCAGCACTTTTAAGCCAGGCTTATGACGCGGATGCTGCGCGAACCAGCACTTGG 1873
 Db 34400 CGGCGCACTACTGAAGTCGCGGATGTGACACCATCGTGCAGGATACCCGAGCGCG 34341
 1874 TTGAGA 1880
 Db 34340 CAGACGA 34334

RESULT 15
 LOCUS BX248337/c 327650 bp DNA linear BCT 11-JUN-2003
 DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 4/14.
 ACCESSION BX248337 BX248333
 VERSION BX248337.1 GI:31617663
 KEYWORDS complete genome.
 SOURCE Mycobacterium bovis subsp. bovis AF2122/97
 ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
 Bacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
 REFERENCE 1
 AUTHORS Garnier,T., Biglmeier,K., Camus,J.-C., Medina,N., Mansoor,H., Pryor,M., Duchoy,S., Grondin,S., Lacroix,C., Monsemppe,C., Simcn,S., Harrie,B., Atkin,R., Doggett,J., Mays,R., Keating,L., Wheeler,P.R., Parkhill,J., Barrall,J.G., Cole,S.T., Gordon,S.V. and Hewinson,G.
 TITLE The complete genome sequence of Mycobacterium bovis
 JOURNAL Online Publication
 REMARK PNAS 10.1073/pnas.1130426100 (Microbiology)
 REFERENCE 2 (baes 1 to 327650)
 AUTHORS Garnier,T.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK, P4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France
 FEATURES
 source 1..327650
 location/Qualifiers
 1..327650
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 698..2170
 gene

CDS

/locus_tag="Mb0912"
 698..2170
 /locus_tag="Mb0912"
 /note="Mb0912", len: 490 aa. Equivalent to Rv0888, len: 490 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 490 aa overlap). Probable exported protein. Equivalent to AK45157.1 from Mycobacterium tuberculosis strain CDC1551 (597 aa) but shorter 17 aa. Contains possible N-terminal signal sequence."

/codon_start=1
 /transl_table=1
 /product="PROBABLE EXPORTED PROTEIN"
 /protein_id="CAD93773.1"
 /db_xref="GI:31617664"

/translation="MDIARITGVQALAVLGVCAVTTAIGSAAPTPSSSTDS
 VDACSPLGASLSALAIIGASVPOVGARQVPGSIPDLALAIIDFLAARNGVPII
 ENRTPVAPNQQVSVPEGGTGVRVPRDAPDGNRTFVRERRGAPGPHOIIVTDQR
 TASFLYADPGFVGTDFPSVANSDTSLHVGAGLYGPFPHDPAVATVAVGNTPT
 DTISGDFSMILVYNIAGLPEPLSAIILPPEYTKELGRKLNAYVAVANEDPAHQFLI
 KSKMPISQTPPEPPTILMPICGVPSSDGLNTSEKRVQRLDQTYECTSDNCLTKGF
 TYSQRLDPSGTDVYNIHHTGTGPTTNALQVANTIQNSAGRAVITGDEFARV
 SDQSALLQFAQVNGLTDAAVQVHGPPTTPPAFCWNGCELDLKIIFRSQGVTL
 QAVSYGNAPRPFNSKGSPLSDHSPAVGVFHYADVAVR"
 complement(2205..3326)
 /gene="cItA"
 /locus_tag="Mb0913c"
 complement(2205..3326)
 /gene="cItA"
 /locus_tag="Mb0913c"
 /EC_number="2.3.3.1"
 /note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c, len: 373 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 373 aa overlap). Probable cItA (alternate gene name: gItA), citrate synthase 2 (EC 4.1.3.7), highly similar to others e.g. CAB95899.1|ALJ59988 putative citrate synthase from Streptomyces coelicolor (387 aa); P39191|CISY BACSU citrate synthase II from Bacillus subtilis (366 aa), FASTA scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa overlap); etc. Also similar to Rv0896|MCY31.24 from Mycobacterium tuberculosis (29.2% identity in 274 aa overlap) and Rv1131. Contains P500480 citrate synthase signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."

gene

CDS

/codon_start=1
 /transl_table=1
 /product="PROBABLE CITRATE SYNTHASE II CITA"
 /protein_id="CAD93774.1"
 /db_xref="GI:31617665"
 /translation="MTVPPENFVFGDGVAFTEIAPDKDGAALYRGVDIDYS
 QRYTFGVNMLLVNDRGSGSLPPEPPLIHSGDVRVDVQGLMPLFVIGVAPLD
 IDDTARQQLARASVMAISVVAQSGRTIYQPAVQRILIDECVTARFMTWQSEPP
 RHIEAIDAYVWASAHEGMNASTFTARVASTGDAVAAISGALGAMGPHGAPAV
 LVPMDEVERRAGDARSVVKGLIDRGEKMGCHRYRAEDPRARVLRAAEELGAPRE
 LPAVDEQALSELREERRPRAIETNVEFMAVVDLPARVPAVNMMPAFETGRTGWCA
 HILEKRLKTLVPSAIVGPRSPESVDGWERVLTAA"
 complement(3413..6061)
 /locus_tag="Mb0914c"
 complement(3413..6061)
 /locus_tag="Mb0914c"
 /note="Mb0914c", len: 882 aa. Equivalent to Rv0890c, len: 882 aa, from Mycobacterium tuberculosis strain H37Rv, (99.9% identity in 882 aa overlap). Probable transcriptional regulatory protein, luxR family, highly similar (but shorter 238 aa in N-terminus) to NP_302202.1|NC_002677 possible transcriptional regulator from Mycobacterium leprae (1106 aa). Also highly similar (generally in part) to others e.g. T50568 probable multi-domain regulatory protein from Streptomyces coelicolor (1334 aa); P10937|NARL ECOLI nitrate/nitrite response regulator protein from Escherichia coli (216 aa), PAS7A scores: opt: 193, E(): 6e-06, (37.4% identity in 99 aa overlap); etc. Also highly similar to others from Mycobacterium tuberculosis e.g. MTCV02B10_22, MTCV08_44,

gene

CDS

MTV036_21, and MTCY31_24. Contains PS00017 ATP/GMP-binding site motif A (P-loop). PS00622 Bacterial regulatory proteins, luxR family signature, and probable helix-turn helix motif from aa 836 to 857 (Score 1559, +4.50 SD). BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL REGULATORS."

```

/codon_start=1
/transl_table=11
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(POBABLE LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:3161766"
/translation="MRALLAQNRVLTLCGTGCGKRLAQTASAEIRDLCEFDVLA
ETLGGIVATARAAGVLPDQGRSTMSLRFGNRRMLWLDNCEHLLACALVAV
ELTGACPELTILATSEPIGMEGITRVPMSITDVAELPADRASVGGRTATV
NAAVGEICRLDGIPLAIEFAARVRSPELIDGDDCCRLLAGVRAGVQOQT
LRASIDSHALLTERBOILFRRLAPVGGFDLAARAVAGSDLPFSLDQTLTLD
KSLVADDCQGRTRRLLETVRVALKEKDGADVAHRHDVYTLAASLNPADN
DHQRLVARETEIDNLRAPAFVSRNGHITELQASSLOPIWFRHARLREGLWNS
ILIEDRFRHASTAVRARALADKAMLTWLTSTPVGATDIIAPQOALAMAREVGP
AALVRLTACGSSGYNAAAPYPAEATDLARATIDDKWTLCOILYMRGVCCTSGD
NALRAAECECDLADITIGDRFVSRHCSIMLSIAQWACNLTEALIELSRETRASN
DVTALLESADARFLLRAQPDQVTHAQMQLALAGDIALAQFANDAVATNGW
RNVALLTRARVATARGBEPLADDAHALACAEHLIYQMPDAMELLAGAGEVSH
SEGVRLGAAALROOTRQVREKIDWAGVATLRAEMGEDPDRAMAGALSTD
EALVAYOGRGERKRPARGMSLTPTEDVAVLVEGSLNMDIAKRLVPSRTQTHL
THVYAKGLASVOLVDEAARSGPS"
/complement(6063..6920)
/locus_tag="Mb0915c"
/locus="Mb0915c"
/complement(6063..6920)
/locus_tag="Mb0915c"
/locus="Mb0915c"
/len=285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
MF_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q1102|MTCY02B10_22
PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in
247 aa overlap); MTV036_21; MTV008_44; MTCY02B10_23. Also
hydrolyases from other organisms."
/codon_start=1
/transl_table=11
/product="POBABLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:3161767"
/translation="MLFNAVHNSLPENIDIDHALIRGEDHPPTCAKCVARGISALGS
LDLRHSLRCVAAAPDVGRCEFPVRVLLANOGLDVRSPLPGTYTLADVEEST
HLMQCPEDMATAIHLDTSEAITNNQGVKRVGSDSFVAAPRASAAACAD
LORTSIADIRLIGLHTGEVOLRDELYGPTINTARLRDLAHGQVYLSATDLYT
GRLPADWLVDLGRHPLRGLRPREWMLCHPDIREFKPLUKAKSSPTSLPQPTT
FVGRAQIS"
/7318..8805
/locus_tag="Mb0916"
/locus="Mb0916"
/7318..8805
/locus_tag="Mb0916"
/locus="Mb0916"
/EC_number="1.14.-.-"
/notes="Mb0916, -, len: 495 aa. Equivalent to Rv0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monooxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monooxygenase from Pseudomonas aeruginosa (491 aa);
CAB59668.1|AL132674 monooxygenase from Streptomyces
coelicolor (519 aa); P12015|CYMO_ACIS cyclohexanone
monooxygenase from Acinetobacter sp. (542 aa); FASTA
scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to Rv0565c, Rv1854c,
Rv1083, etc from Mycobacterium tuberculosis. Has

```

```

hydrophobic stretch at N-terminus."
/codon_start=1
/transl_table=11
/product="PROBABLE MONOOXYGENASE"
/protein_id="CAD93777.1"
/db_xref="GI:31617668"
/translation="MTCRCPTVAIVGAGMCMVAITLSAGITDVCITYEKADVGGT
WRDTPGLTCDVPSRLYQYSPKRNPMWTFESRGGTIDYLRGILRERGLRIRFG
ATVVSARFDGRWLRLTDSGTSTVPLISATSVLHPRIPLIAGIDFRTVFSAR
VDFHAPPLGLSLAVKAYSLSFETFAVALSNPGLRKLVGACRASLRVDPRLRAL
TPDYEPWCKLVMGSGFYRAIORDDVELVAGIDVHNRGIVDDGLVHDVIVLAT
GPSHAFERFMOLTGRRGRIIDVWODGPHAHQTVAILPGPNPFMMGLGPGSVCNPL
TVAESQAEHIVQIMKMRGEPDTPKSAITBATNTVIRAMPNTVTTGGDSWTL

```

```

Query Match      14.7% Score 312.6; DB 1; Length 327650;
Beet Local Similarity 54.0%; Pred. No. 3.3e-60;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

QY 572 ACTGGTTTGGACCCCTGACAGCTTCACTTTCTTGGAATGAACCTCCCAATATGACACC 631
DB 21614 ACGCGGTGCTAGACCGGGGATCTTTCTGTACCTGGGATAGACCGCTGGCGGTGCGG
QY 632 TCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGATGAAT 691
DB 21554 TAGCCGACTCCTATGACCGGGAGCTGCGCCCGCTCGGGGGGACACCGCGCGAGCAAT 21495
QY 692 CGGTAAATTACTGAGAAAGCAACCGTGAAGGCAATCCGTAGACCGTATTTTGTCCGATT 751
DB 21494 CGGTGACAGCCGTGAGAGGAGCGGTATTCGGGGCGGGGTGGCGGTGTGAGAGT 21435
QY 752 TTTCCTTCCTCGGGGTTCTTTTGGGACCGGTGCGCGTGGCGCATCATGAAGCGCATTC 811
DB 21434 TCGACTTCTGGGGGCTTCATTGGGGTGGACACGGCCGAAGATACCGCGCGGTG 21375
QY 812 ACGCGGCAAGAGCTGAAATCCCACTGCTGCTCCCTGCTCCGTTCCGGTGGCGCGCA 871
DB 21374 AGCGGGGACCGCGAGCGCGCTCGCTACCGGTGACCAACTCGGAGGACACCGCA 21315
QY 872 TGCAGAAAGCAATCGACTTTTGTATGATGATGATGATCAACCGCGCGCTGTGACGCTC 931
DB 21314 TGCAGAAAGCAAGCTGCTGCTTTCTCAGATGATGAAATCGCTGCGGCATCACTGC 21255
QY 932 ACGCGGAGGCGCATTTGCGCTTCCTGATTTTGGCAATCCAGATGGGTGGCGCA 991
DB 21254 ACAACCAAGCGCGCTGCTTACCTGATCTATTTGCGCAATCCAGACAGCGGTGAGTTT 21195
QY 992 TGGCTTGTGGGTTTCATCTGGGCATCTCACTTTTGCAGAACCCGCGCGAGATAGTTT 1051
DB 21194 TCGCGTGTGGGCTCGCTGGGGCATCTCACCGCTGCCAGCGGGCGCTGATGGCT 21135
QY 1052 TCGTGGTCTCTCGCGGTGTGAGTTTAAACACTGGGATCGCTTCCAGACGTTGTCAGC 1111
DB 21134 TTCTGGGACCAAGGCTCTATGAGTTCTCTATGCGACACCCCTTCCATCCGCGTCCAAA 21075
QY 1112 AGCGGAGAAATTTGTGAAAACCTGATGATGATGATGATGATGATGATGATGATGATG 1171
DB 21074 CGCGCGAATCTACGCGCGCATGAGATATGACGCGCTCGTTGACATGACCGGCTAC 21015
QY 1172 GTGACGCGTGGCAAAACCTTCAGGTTATTCAGCGGTAGAGGCAAGATCGTTT 1231
DB 21014 GACCGATGCTGATCGTGGTTGACGCTCATGACGCTCCGAAACCGCTTCCGGCAC 20955
QY 1232 CTCACCAACTCTGCGGTGCGA---CTTCGCGATGAGAGCGATTTGCGCTTCTGTC 1288
DB 20954 CGAGACGCGCGCGCTGATCCGATGTCACAGGTGGACCTCGGTGTGAGATCGCGCC 20895
QY 1289 ACCCGAGAGGCGCTGATTCGGGAGATATATGAAACGTTGGGGGACAGCTGTCAGAC 1348
DB 20894 GCGCGGACCGCGCGCTGATTCAGGACCTACTGCGACAGGCGCACGACGCGGTGTTT 20835
QY 1349 TTTCGTGCGCGGTGCGCATTTGAGCGCGGCTGTGCGCGCTTGGCGCGCATG 1408

```

```
Db 20834 TGTCAAGAACCGATCAAGGCGAAGCG---GCGACCAAGCTGTGCGCTGCGCCGCTTTG 20778
QY 1409 GGGGCGGGCCCGGNGTGTGATTTGGGCG-----GATGCGCGCTTACAGC 1453
Db 20777 GCGGCGAACCCACGGGTGCTCTCGGCGACAAAGGCGAGTAGGCGGGGGAGACATG 20718
QY 1454 TTGGCCCGAGAGAGCTGCGTTTTTGGCGCTGTGCGCATTTGCTGCGCGCGAGCTAAAC 1513
Db 20717 TCGGGCCCGCTCGTTACGCGAAGCCGAGCGGGATGCGGCTCGCGCGAGCTGTGCC 20658
QY 1514 TCCCGATCGTGTTCATCATGACACCTCCGCGCGCGAATTGTGCAAGCGCGCTGAGAGC 1573
Db 20657 TCCCGTGTGTGTGATTTGACGCGCGGACCCGCGTTGTGCGCGCGACCGAACAGG 20598
QY 1574 TCGGATCGCAAGCTGTGATTTGGCGGACCTTGTCCAAAGTTATCGAGCTCCCTCCCA 1633
Db 20597 GCGGCTGCGCGGCGACATGCGGCAATTGCGCGAGCTCGTCAAGCTGATACCCCGA 20538
QY 1634 CCGTTTCGGTCAATTATTGCTCAGGGCGTTGGCGGTGCGCGCTGGCATGTGCCGCG 1693
Db 20537 CCGTGTGATCTGTGCTGGGCGAGGGGAGCGGCGCGCGCTGGGATGTGCCGCG 20478
QY 1694 ATCTGCTTACGCGGCGGAAACGCGTGTGTCTGCGATTGCAACGAGGCGCTCGG 1753
Db 20477 ACCGGGTGTGCGCGCATCTCACGCGTGTGCGCGCTTGTCTCCGAAAGAGCAAGC 20418
QY 1754 CCATCTCTTCCGCGACACCAACGCGCGCGGAAATCATAGAGCAAGGCGTGCAG 1813
Db 20417 CGATGTGTTCGAGACATGCTCTCATGCGCGCAACTGCTGCCGCCCAAGGCATCCGT 20358
QY 1814 CGCAGCACTTTTAAGCCAAAGGCTTATGACGCGATGTCGCCGAAACCGAGCACTTG 1873
Db 20357 CCGCGGACCTACTGAAGTGGGGATTTGTGACACCATGTGCGCGAGTACCCCGAGCGCG 20298
QY 1874 TTGAAGA 1880
Db 20297 CAGACGA 20291
```

Search completed: November 13, 2003, 00:10:28
Job time : 7983.97 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 15:05:16 ; Search time 588.017 Seconds
(without alignments)
9746.164 Million cell updates/sec

Title: US-10-024-370-1

Perfect score: 2123

Sequence: 1 ctccgagcgagctgcgtgatc.....gaccttcacgtcgtgattc 2123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_190un03.*
1: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	2123	22	AAC82732
2	2111	99.4	349980	22	AAH68526
3	1575	74.2	1575	22	AAH67755
4	1473	69.4	1473	22	AAH65894
5	1473	69.4	1473	22	AAC82733
6	312.6	14.7	1488	22	AAH51981
7	312.6	14.7	4403765	22	AAI96683
8	312.6	14.7	4411529	22	AAI96682

9	168	7.9	168	25	ACA00248
10	131	6.2	354	22	AAH68407
11	122.8	5.8	873	23	AAS54163
12	122.2	5.8	876	25	ABZ39000
13	121	5.7	867	23	AAS53191
14	121	5.7	6021	20	AAH13395
15	121	5.7	6021	24	ABG99190
16	119.2	5.6	676	24	ABX65750
17	119.2	5.6	877	23	AAS5651
18	119.2	5.6	877	24	ABX65721
19	119.2	5.6	1053	19	AAH14153
20	118.6	5.6	349980	21	AAH21608
21	118.6	5.6	1437668	21	AAH14490
22	108.6	5.1	23673	23	ABZ75344
23	108	5.1	993	23	AAS56217
24	102.8	4.8	915	23	AAS52496
25	99.8	4.7	3118	23	AAS88487
26	99.8	4.7	3118	23	AAS89710
27	97.8	4.6	867	23	AAS55883
28	96.8	4.6	654	24	ABK78240
29	96.2	4.5	864	25	ABX06060
30	96.2	4.5	867	22	AAH90763
31	96.2	4.5	891	23	AAS53478
32	96.2	4.5	1595	19	AAZ96348
33	96.2	4.5	19702	19	AAV52140
34	96.2	4.5	1810121	17	AAT42063
35	96.2	4.5	2162598	25	ABG56454
36	94.2	4.4	873	24	ABN68650
37	94.2	4.4	2155561	24	ABN71527
38	93.6	4.4	495269	24	ABQ67195
39	93.6	4.4	3011208	24	ABO69245
40	93	4.4	867	23	AAS5630
41	93	4.4	891	22	AAH90875
42	92.8	4.4	1001	22	AAH91440
43	92.8	4.4	1001	22	ABK37820
44	91	4.3	2365589	24	ABA90521
45	88	4.1	2944528	24	ABA03041

ALIGNMENTS

RESULT 1	
AAC82732	standard: DNA; 2123 BP.
AC	AAC82732;
XX	
DT	15-MAR-2001 (first entry)
DE	C. glutamicum accDA DNA.
XX	
KW	L-amino acid; accDA; coryneform microorganism; L-lysine; animal feed; medicine; pharmaceutical industry; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
FN	EP1055725-A2.
XX	
PD	29-NOV-2000.
XX	
PF	10-MAY-2000; 2000EP-0109842.
XX	
PR	27-MAY-1999; 99DE-1024365.
XX	
PA	(DEGS) DEGUSA-HUELS AG.
XX	(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
PI	Tilg Y, Eggeling L, Birkmann B, Sahn H, Moeckel B,
XX	WPI, 2001-042411/06.
DR	P-Psdb; AAB45789.
XX	

C. glutamicum deri
C glutamicum codin
Pseudomonas aerugi
N. gonorrhoeae nuc
Enterococcus faeca
Enterococcus faeca
Enterococcus faeca
Helicobacter pylor
Helicobacter pylor
H. pylori GHP0 451
Neisseria meningit
N. meningitidis B
Human R128 gene C
Salmonella typhi D
E. coli DNA for ce
DNA encoding novel
DNA encoding novel
Streptococcus pneu
Bacillus clausii g
S. pneumoniae type
CFE 67 coding sequ
Haemophilus influe
S. pneumoniae deri
Streptococcus pneu
Haemophilus influe
Streptococcus pneu
Streptococcus poly
Listeria innocua c
Listeria innocua D
Streptococcus pneu
2CFE 67 coding seq
Haemophilus influe
DNA sequence upstr
Genomic sequence o
Listeria monocytog


```

Db      1801  CAAAGCGTCAGAGCGCAGCATTAAAGCAAGGGCTTATCGACGGATCGCGCAA 1860
Qy      1861  ACCGAGCATTTGTTGAAGAAATTCGCGCAAAATAGCAAGCCCTCCCAATTGAT 1920
Db      1861  ACCGAGCATTTGTTGAAGAAATTCGCGCAAAATAGCAAGCCCTCCCAATTGAT 1920
Qy      1921  AACCAATCCGAGAGGGCGGAGCGCAGCAGTCCCTTACACAGATTTAGCGTGA 1980
Db      1921  AACCAATCCGAGAGGGCGGAGCGCAGCAGTCCCTTACACAGATTTAGCGTGA 1980
Qy      1981  TAAAGAAATTTATGCGCTGATCAATCGATGAAACACAGGTTACGGCCAGACGTG 2040
Db      1981  TAAAGAAATTTATGCGCTGATCAATCGATGAAACACAGGTTACGGCCAGACGTG 2040
Qy      2041  GTGGCGGAGACCCCTCAGAGGCGCTTGAAGCAGCTTGGCGGAATGCTGACGAT 2100
Db      2041  GTGGCGGAGACCCCTCAGAGGCGCTTGAAGCAGCTTGGCGGAATGCTGACGAT 2100
Qy      2101  GCCGACCTTCATGCTGGAATTC 2123
Db      2101  GCCGACCTTCATGCTGGAATTC 2123

```

RESULT 2
AAH68526/c
ID AAH68526 standard; DNA; 349980 BP.

AAH68526;
26-SEP-2001 (first entry)
C glutamicum coding sequence fragment SEQ ID NO: 7061.

Coriobacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis; ds.
Coriobacterium glutamicum.

EP108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOWA) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

Novel polynucleotides derived from Coriobacterium bacteria, for identifying
mutation point of a gene, measuring expression of a gene, analyzing
expression profile or pattern of a gene and identifying homologous gene
Disclosure; SEQ ID NO: 7061; 246bp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein
sequences from the Coriobacterium bacterium Coriobacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of coriobacterium bacterium, measuring expression amount and
analysing the expression profile or expression pattern of a gene derived
from Coriobacterium bacterium, and identifying a homologue of a gene derived
from coriobacterium bacterium. Coriobacterium bacteria are useful for producing
amino acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present sequence is a nucleic acid described
in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 349980 BP; 79274 A; 90638 C; 98727 G; 81341 T; 0 other;

Query Match 99.4%; Score 2111; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Qy      1  CTGAGCGGAGTCGGTATCGGCCACTCTCTAAGCAATGCCGCTTTAAATAAGCA 60
Db      281620  CTGAGCGGAGTCGGTATCGGCCACTCTCTAAGCAATGCCGCTTTAAATAAGCA 281561
Qy      61  CTTATATGTTTCTACACATCTGGCCGACCAACGAAGTATGTTTCATCAAGCTA 120
Db      281560  CTTATATGTTTCTACACATCTGGCCGACCAACGAAGTATGTTTCATCAAGCTA 281501
Qy      121  AACGTGTAATGTGAAGTTAAGTAACTTAATGCAATGCAATGCGATGCAATTTGAAATCTC 180
Db      281500  AACGTGTAATGTGAAGTTAAGTAACTTAATGCAATGCAATGCGATGCAATTTGAAATCTC 281441
Qy      181  ACTCCGCCAATATCTTAATTAATTAAGTATGTTTACCTGCAATTTATTAAGT 240
Db      281440  ACTCCGCCAATATCTTAATTAATTAAGTATGTTTACCTGCAATTTATTAAGT 281381
Qy      241  TCCCGATCAATCCCTCTTTTACCCGAAATACCTTTTGCAGAAATGCAACACACA 300
Db      281380  TCCCGATCAATCCCTCTTTTACCCGAAATACCTTTTGCAGAAATGCAACACACA 281321
Qy      301  GTGCAATAGTTAACGGGCTTCAACAGCAATCTGTCGGTTTATAGCTATGTTTCCGG 360
Db      281320  GTGCAATAGTTAACGGGCTTCAACAGCAATCTGTCGGTTTATAGCTATGTTTCCGG 281261
Qy      361  ACGTCTAGCAAAAAGTATGTTTGTAGATGAAGAACGATTAATCCGTTATTTTACGCA 420
Db      281260  ACGTCTAGGC-AAAAGTATGTTTGTAGATGAAGAACGATTAATCCGTTATTTTACGCA 281202
Qy      421  TCGATAGCTTAATTTGGGCTTATGATCTTCGCTTAAATATGATGCAAGACATTGCA 480
Db      281201  TCGATAGCTTAATTTGGGCTTATGATCTTCGCTTAAATATGATGCAAGACATTGCA 281142
Qy      481  ATTAATTAACAAAGCATTTTTCGCGCGTGGAGAACGTTTTCGACTATGCTGGGGC 540
Db      281141  ATTAATTAACAAAGCATTTTTCGCGCGTGGAGAACGTTTTCGACTATGCTGGGGC 281082
Qy      541  ATGGAACACACTTCAGATTGACGCTCATAGACTGTTTGAACCTGACAGCTTCATT 600
Db      281081  ATGGAACACACTTCAGATTGACGCTCATAGACTGTTTGAACCTGACAGCTTCATT 281022
Qy      601  TCTTGAATGAATCTCCCAATATGACAACTCAATCAAGCTATGCAAGACCTTGGAG 660
Db      281021  TCTTGAATGAATCTCCCAATATGACAACTCAATCAAGCTATGCAAGACCTTGGAG 280962
Qy      661  CGGGCTCGAAGCAAGGCAAAATGCGATGATCGGTAATTAATCGAGAGAGGACCGTGGAG 720
Db      280961  CGGGCTCGAAGCAAGGCAAAATGCGATGATCGGTAATTAATCGAGAGAGGACCGTGGAG 280902
Qy      721  GGCATTCGCGTAGCGCTTATTTTGTCCGATTTTCTTCTCGCGGCTTCTTTGGGACAG 780
Db      280901  GGCATTCGCGTAGCGCTTATTTTGTCCGATTTTCTTCTCGCGGCTTCTTTGGGACAG 280842
Qy      781  GTCCGCTGCGTGGCGCATATGAAGCGATTACCGCGCACAGAGCTGAATCTCCACTG 840
Db      280841  GTCCGCTGCGTGGCGCATATGAAGCGATTACCGCGCACAGAGCTGAATCTCCACTG 280782
Qy      841  CTGGCTCCCTGCTGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db      280781  CTGGCTCCCTGCTGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 280722
Qy      901  ATGGTGTTCATTAACCGGCTGTGCAAGGTCACCGGAGGCGCATTTGCTTCTGCTG 960
Db      280721  ATGGTGTTCATTAACCGGCTGTGCAAGGTCACCGGAGGCGCATTTGCTTCTGCTG 280662

```

QY 961 TATTGGCGCAATCCACGATGGGTGGCGGCATGGGCTGTGGGGTTCAATCTGGCATCTC 1020
Db 280661 TATTGGCGCAATCCACGATGGGTGGCGGCATGGGCTGTGGGGTTCAATCTGGCATCTC 280602
QY 1021 ACTTTTGGGGAACCCGGCGCGAGATAGGTTTCTGTGGTCTCGCGGTGTGGAGTTAAC 1080
Db 280601 ACTTTTGGGGAACCCGGCGCGAGATAGGTTTCTGTGGTCTCGCGGTGTGGAGTTAAC 280542
QY 1081 ACTGGGCAATGGGCTTCCAGACGGTGTGACAGCGGAGAAATTTGTGAAAATCTGTGTG 1140
Db 280541 ACTGGGCAATGGGCTTCCAGACGGTGTGACAGCGGAGAAATTTGTGAAAATCTGTGTG 280482
QY 1141 ATTGATGGAATTTGTGTCGCACTCCAAATTTGCGTGGAGGGGAAAAAACCCTCAAGTT 1200
Db 280481 ATTGATGGAATTTGTGTCGCACTCCAAATTTGCGTGGAGGGGAAAAAACCCTCAAGTT 280422
QY 1201 ATTCAGCCGGTATGAGCAACGAGATGTTTCTCCAAACAATCTGGCGTGGCACTTCCG 1260
Db 280421 ATTCAGCCGGTATGAGCAACGAGATGTTTCTCCAAACAATCTGGCGTGGCACTTCCG 280362
QY 1261 GTGATGAGGCGATTTGCGGTTCTGTGACCCGCAAGGCTTGAATGGGGAATTATG 1320
Db 280361 GTGATGAGGCGATTTGCGGTTCTGTGACCCGCAAGGCTTGAATGGGGAATTATG 280302
QY 1321 GAAACGTTGGGGGAGAGGTGTCAAGCTTCTGGTGGCGGCTGGGCGCAATTAGCCCG 1380
Db 280301 GAAACGTTGGGGGAGAGGTGTCAAGCTTCTGGTGGCGGCTGGGCGCAATTAGCCCG 280242
QY 1381 GCTGTGCGCGTGTGCCCTGTGCGCGCATCGGGGCGCGGCCGTGTGTCGATTTGGCGAGAT 1440
Db 280241 GCTGTGCGCGTGTGCCCTGTGCGCGCATCGGGGCGCGGCCGTGTGTCGATTTGGCGAGAT 280182
QY 1441 CGCCGCTTACGCTTTGGGCGCGAGAGCTGCGTTTGGCGGTGTGGCATTTCCGTGCGG 1500
Db 280181 CGCCGCTTACGCTTTGGGCGCGAGAGCTGCGTTTGGCGGTGTGGCATTTCCGTGCGG 280122
QY 1501 CGCGAGCTAAACCTGCGCATGTGTCCATTCGACACCTCCGGGCGCGAATTTGTCGAG 1560
Db 280121 CGCGAGCTAAACCTGCGCATGTGTCCATTCGACACCTCCGGGCGCGAATTTGTCGAG 280062
QY 1561 GCGGCTGAGAGCTTGGCATGTGCAATTCGACACCTTGTCCAGCTTAATCGAC 1620
Db 280061 GCGGCTGAGAGCTTGGCATGTGCAATTCGACACCTTGTCCAGCTTAATCGAC 280002
QY 1621 GCTCCCTTCCCAACGTTTCCGTCAATTTGTCAGGCGGTTGGCGGCGCTGCGC 1680
Db 280001 GCTCCCTTCCCAACGTTTCCGTCAATTTGTCAGGCGGTTGGCGGCGCTGCGC 279942
QY 1681 ATGCTGCGCGCGATCTGTCTGACGCGGCGGAAAAAGGTGGTGTCCGCAATTGCAACA 1740
Db 279941 ATGCTGCGCGCGATCTGTCTGACGCGGCGGAAAAAGGTGGTGTCCGCAATTGCAACA 279882
QY 1741 GAGGCGCGCTCGCCCATCTTCCGCGACAACAACAACGCGCGGGAATATAGAGCGA 1800
Db 279881 GAGGCGCGCTCGCCCATCTTCCGCGACAACAACAACGCGCGGGAATATAGAGCGA 279822
QY 1801 CAAAGCGTGCAGGCGCGACGCACTTTAAACCAAGGGCTTAATGACGGATGTGCGCGAA 1860
Db 279821 CAAAGCGTGCAGGCGCGACGCACTTTAAACCAAGGGCTTAATGACGGATGTGCGCGAA 279762
QY 1861 ACCGAGCACTTTGTTGAAGAAATTTCTGGGCAACATACACAAGCCCTCCGGAATTTGAT 1920
Db 279761 ACCGAGCACTTTGTTGAAGAAATTTCTGGGCAACATACACAAGCCCTCCGGAATTTGAT 279702
QY 1921 AACCAATCCGAGAGGCGCGAGCGCAAGTGTGTCACACGATTTGAGCGTTTACGCGAG 1980
Db 279701 AACCAATCCGAGAGGCGCGAGCGCAAGTGTGTCACACGATTTGAGCGTTTACGCGAG 279642
QY 1981 TAAAGAAAATTAATGCGTGTATCAAAATCGATATGAAACAACAGGGTACGGCCAGACAGTGG 2040
Db 279641 TAAAGAAAATTAATGCGTGTATCAAAATCGATATGAAACAACAGGGTACGGCCAGACAGTGG 279582

QY 2041 GTGGCGGAAACCCCTCAGGCGCGTAAGCAAGCTCTGGCGGATGTGATGACGACGCTCC 2100
Db 279581 GTGGCGGAAACCCCTCAGGCGCGTAAGCAAGCTCTGGCGGATGTGATGACGACGCTCC 279522
QY 2101 GCCGACCTTCATGCTCGTAATTC 2123
Db 279521 GCCGACCTTCATGCTCGTAATTC 279499

RESULT 3
AA67755
ID AA67755 standard; DNA; 1575 BP.
XX
AC AA67755;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:25.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN M0200100805-A2.
PD
XX 04-JAN-2001.
XX
PE 23-JUN-2000; 2000MO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BAD) BASF AG.
XX
XX Pompeius M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

DR MPI; 2001-071486/08.
DR P-PSDB; AAB76522.

Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
transformation -

Claim 3; Page 173-175; 1119pp; English.

AAAF7743 to AAFF6808 encode the Corynebacterium glutamicum membrane
construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAFF6808 and AAFF6802 represent sequencing primers which are used in an
CC example from the present invention.

Sequence 1575 BP; 309 A; 434 C; 474 G; 358 T; 0 other;

Query Match 74.2%; Score 1575; DB 22; Length 1575;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 429 CTAATTTGGGCTTAAATCTTCCGCTTAATATGATGACAGACATTGCAATTAATTA 488
DB 1 CTAATTTGGGCTTAAATCTTCCGCTTAATATGATGACAGACATTGCAATTAATTA 60
QY 489 ACAAGGCCATTTTGGCGCCGAGAGAGGTTTCCGACATATGATGAGGAGATGAACA 548
DB 61 ACAAGGCCATTTTGGCGCCGAGAGAGGTTTCCGACATATGATGAGGAGATGAACA 120
QY 549 CACTTCAGATTCAGCTCCTGATGACTCGGTTTGGACCTGACAGCTTCATTCTTGAA 608
DB 121 CACTTCAGATTCAGCTCCTGATGACTCGGTTTGGACCTGACAGCTTCATTCTTGAA 180
QY 609 TGAATCTCCCAATATGACAACTCAATTAAGGCTATGACAGACCTTGGAGCGGCTCG 668
DB 181 TGAATCTCCCAATATGACAACTCAATTAAGGCTATGACAGACCTTGGAGCGGCTCG 240
QY 669 AAGCAGGCGCAATGCGATGATCGTATTAATCTGAGAGAGGACCGTGGAGGAGATTC 728
DB 241 AAGCAGGCGCAATGCGATGATCGTATTAATCTGAGAGAGGACCGTGGAGGAGATTC 300
QY 729 GGTAACCGTTATTTTGTCCGATTTTCTTCTCGCGGCTTCTTTGGGACGCTCGCTC 788
DB 301 GGTAACCGTTATTTTGTCCGATTTTCTTCTCGCGGCTTCTTTGGGACGCTCGCTC 360
QY 789 GGTGGCATCATGAAGGCGATTACCGCGCCACAGAGCTGAAATCTCCACTGCTGTC 848
DB 361 GGTGGCATCATGAAGGCGATTACCGCGCCACAGAGCTGAAATCTCCACTGCTGTC 420
QY 849 CCTCTCTCCGCTGGTGGCGCGCATGACAGAGAACATGAGCTTTTGTGATGATGCTGC 908
DB 421 CCTCTCTCCGCTGGTGGCGCGCATGACAGAGAACATGAGCTTTTGTGATGATGCTGC 480
QY 909 CATTAACCGCGCTGTGACAGCGTCAACCGGAGGCGCATTTGCCGTTCTGTGTATTTGG 968
DB 481 CATTAACCGCGCTGTGACAGCGTCAACCGGAGGCGCATTTGCCGTTCTGTGTATTTGG 540
QY 969 CATTCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1028
DB 541 CATTCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 1029 GGAACCCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088
DB 601 GGAACCCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 1089 TGCGCTTCAGACGGTGTGACAGACGCGAGAAATTTGTGAAACTGTGTGATGATGATG 1148

```

```

DB 661 TGCGCTTCAGACGGTGTGACAGACGCGAGAAATTTGTGAAACTGTGTGATGATGATG 720
QY 1149 AATTGTGTGCGCATCTCCCAATTTGCGTGCAGCGGTGGCAAAAACCTTCAAGTTATTCAGCC 1208
DB 721 AATTGTGTGCGCATCTCCCAATTTGCGTGCAGCGGTGGCAAAAACCTTCAAGTTATTCAGCC 780
QY 1209 GGTAGAGGCAACGATGCTTTTCTCCAACTCTGAGCGTGGACCTTCCGGTATGGA 1268
DB 781 GGTAGAGGCAACGATGCTTTTCTCCAACTCTGAGCGTGGACCTTCCGGTATGGA 840
QY 1269 GCGCATTTGGCGCTTCTGTGACCCGAGAGGCTGGAATTCGGAGATTAATGAAACGTT 1328
DB 841 GCGCATTTGGCGCTTCTGTGACCCGAGAGGCTGGAATTCGGAGATTAATGAAACGTT 900
QY 1329 GGGGGGCAACGCTTCGCAAGCTTTCTGTGCGGTGCTGTGCGATGAGCCGCGCTGTGCG 1388
DB 901 GGGGGGCAACGCTTCGCAAGCTTTCTGTGCGGTGCTGTGCGATGAGCCGCGCTGTGCG 960
QY 1389 CGTTGCCCTGGGCGCATTCGGGGGCGGCGCGGTGGTGTGATTTGGGAGAGATCGCGCTT 1448
DB 961 CGTTGCCCTGGGCGCATTCGGGGGCGGCGCGGTGGTGTGATTTGGGAGAGATCGCGCTT 1020
QY 1449 CACGCTTGGGCGCGAGAGCTGCTGTTTGGCGGTGCTGAGCAATTTGCTGCGCGAGCT 1508
DB 1021 CACGCTTGGGCGCGAGAGCTGCTGTTTGGCGGTGCTGAGCAATTTGCTGCGCGAGCT 1080
QY 1509 AACTCTGCGCATCTGTGTCATGATGACACCTTCGCGCGCGCAATTTGTCAGAGCGGCTGA 1568
DB 1081 AACTCTGCGCATCTGTGTCATGATGACACCTTCGCGCGCGCAATTTGTCAGAGCGGCTGA 1140
QY 1569 GGAGCTGCGCATTCGCAAGCTGATGAGCGGCGGACCTTGTCCAAGCTTATCGACGCTCCCT 1628
DB 1141 GGAGCTGCGCATTCGCAAGCTGATGAGCGGCGGACCTTGTCCAAGCTTATCGACGCTCCCT 1200
QY 1629 CCCACCGTTTCGATCATTAATTTGTCAGAGGCGTGGCGGCTGAGCATGCTGCC 1688
DB 1201 CCCACCGTTTCGATCATTAATTTGTCAGAGGCGTGGCGGCTGAGCATGCTGCC 1260
QY 1689 CGCGCATCTGTGCTACGCGGCGGAAACGCGTGTGCTGCGCATTTGCCACAGAGGCGC 1748
DB 1261 CGCGCATCTGTGCTACGCGGCGGAAACGCGTGTGCTGCGCATTTGCCACAGAGGCGC 1320
QY 1749 CTGGCGCATCTGCTTCCGCGCACCAACACAGCGCGGAAATCATGAGCGCAAGAGGCT 1808
DB 1321 CTGGCGCATCTGCTTCCGCGCACCAACACAGCGCGGAAATCATGAGCGCAAGAGGCT 1380
QY 1809 GCAAGCGCACGACATTTTAACCAAGGCTTATGACAGGATGCTGCGCGAAACCGAGCA 1868
DB 1381 GCAAGCGCACGACATTTTAACCAAGGCTTATGACAGGATGCTGCGCGAAACCGAGCA 1440
QY 1869 CTTTGTGAAGAAATTTCTCGGCAATGACAGCGCTTCTCCGATTTGATTAACAATTC 1928
DB 1441 CTTTGTGAAGAAATTTCTCGGCAATGACAGCGCTTCTCCGATTTGATTAACAATTC 1500
QY 1929 GGAAGGCGGAGGAGCGGACAGTGGCTTCAACGATTTGAGGCTTGAAGCGATAAGAA 1988
DB 1501 GGAAGGCGGAGGAGCGGACAGTGGCTTCAACGATTTGAGGCTTGAAGCGATAAGAA 1560
QY 1989 ATTATGCGCTGATCA 2003
DB 1561 ATTATGCGCTGATCA 1575

```

RESULT 4
AAH65894
ID AAH65894 standard; DNA; 1473 BP.

AAH65894;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 929.


```

Db      1261 GACACCAACGACGCCGGAATCATAGAGCAGAAAGGCTGACGCGACCACTTTTA 1320
Qy      1828 AGCCAAAGGCTTATGACGGGATTCGTCCGCCAAACCGACACTTTGTTGAATAATTCTC 1887
Db      1321 AGCCAAAGGCTTATGACGGGATTCGTCCGCCAAACCGACACTTTGTTGAATAATTCTC 1380
Qy      1888 GGCACATCAGCAAGCCCTCTCCGAATTGATTAACAATCCGAGAGGGGCGGACCGCAGC 1947
Db      1381 GGCACATCAGCAAGCCCTCTCCGAATTGATTAACAATCCGAGAGGGGCGGACCGCAGC 1440
Qy      1948 AGTCGCTTACACGATTTGAGCGTTTAAAGCGAG 1980
Db      1441 AGTCGCTTACACGATTTGAGCGTTTAAAGCGAG 1473

```

RESULT 6
AAH51981
ID AAH51981 standard; DNA; 1488 BP.

```

AC      AAH51981;
XX      DT      04-SEP-2001 (first entry)
XX      DE      Mycobacterium tuberculosis potential drug target gene SEQ ID 35.
XX      KM      Drug target; growth; organism viability; characterisation; de.
XX      OS      Mycobacterium tuberculosis.
XX      PN      WO200135317-A1.
XX      PD      17-MAY-2001.
XX      PF      13-NOV-2000; 2000MO-US31152.
XX      PR      12-NOV-1999; 99US-0165086.
XX      PR      12-NOV-1999; 99US-0165124.
XX      PR      01-FEB-2000; 2000US-0179531.
XX      PA      (BECG) UNITV CALIFORNIA.
XX      PI      Eisenberg D, Rotstein SH, Marcotte EM;
XX      DR      WPI: 2001-329193/34.
XX      DR      P-PSDB; AAG81130.

```

Identifying nucleotide or polypeptide sequence for use as drug target, involves providing algorithm that analyzes a functional relationship between nucleotide or polypeptide sequences, and comparing the sequences -

Disclosure; Page 74; 207pp; English.

This invention relates to a method for identifying a nucleotide or polypeptide sequence that may be a drug target, or essential for growth or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092 represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium tuberculosis proteins which are potential drug targets. The DNA and protein sequences are used to illustrate the method of the invention. The method involves providing an unknown nucleotide or polypeptide sequences, and comparing it to a number of sequences along with at least one algorithm capable of analysing a functional relationship between nucleotide and polypeptide sequences. The method is useful for characterising the function of nucleic acids and polypeptides that may be useful as a target for a drug or essential for the growth or viability of an organism.

Sequence 1488 BP; 220 A; 509 C; 524 G; 235 T; 0 other;

Query Match 14.7%; Score 312.6; DB 22; Length 1488;
Best Local Similarity 54.0%; Pred. No. 1.1e-84;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```

Qy      572 ACTCGTTTGGACCTGACAGCTTCATTTCTTGAATGAAGCTCCCAATATGACAACC 631
Db      32 ACAGCGTGTCTAGACCCGGGGAATCTTTGCTACAGTGGATTAAGCAAGCCGCTGGCGG 91
Qy      632 TCAATCAAGGCTATGACAGACCTTGAAGCGGGCTCGAAGCAAGCCAAATGCGATGAAT 691
Db      92 TAGCCGACTCTATGCGCGGAGCTGGCCCGCTCGGGGCGGACACCGGCGGACGAAT 151
Qy      692 CGGTATTAATCTGAGGAAGGACACGTGAGAGGCATTCGGGTAGCCGTATTTTGCCGATT 751
Db      152 CGGTGACAGCCGCTGAGAGGACGCGTATTCGAGCGGCGGCTGGCTGTGAGT 211
Qy      752 TTTCCTTCCTCGAGGCTTCTTTTGGGACAGGTCGCGTGGTGGCGCATGATGAAGGCGATT 811
Db      212 TCGACTTCTGAGCGGCTCGCATTTGGGTGGACGCGCCGAACGATACCGCCGCGCTCG 271
Qy      812 ACCGCGCCACAGACTGAAATCCCACTGCTGTCTCCCTGCTTCGGTGTGCGGCA 871
Db      272 AGCGGGGCGACCGCGGAGCGGCTGCGCTACTGAGGCTCACCAGCTCGGAGGCACTCCGCA 331
Qy      872 TGCAGAAAGACATGACGCTTTTGTCAATGATGCTCATTAACCGCGGCTGTGAGCTC 931
Db      332 TGCAGAAAGGACGCGTGGCTTCTGCAAGATGTAAGATCGCTGCGGCATCCAGCTGC 391
Qy      932 ACCGCGAGCGCATTTGCGCTGCTGTGATTTTGGCATATCCAGATGAGGCGGCA 991
Db      392 ACAACAGGCGCGCTGCTGCTTACTGATTTTGGCCATCCGACCAAGGATGAGTTT 451
Qy      992 TGAGCTGTGGGATTCATCTGGGCAATCACTTTTGGGAAACCGGCGCGCAATAGATT 1051
Db      452 TCGCGTGTGGGAGCTTCGCTGGGAGCATCTCACCGTCCGAGCGGGCGCTGATCGGCT 511
Qy      1052 TCTTGGGTCTCGGTGTGATTAACACTGGGCAATGGCGTTCCAGAGGTGTGAGC 1111
Db      512 TTTGAGGACCAAGGCTATGATGCTATAGCGACCCCTTCCATCCGCGCTCCAAA 571
Qy      1112 AGCGGAGAAATTTGTGTAACCTGTGTGATGATGATTTGTGTCCTCCACTCCAAATTGC 1171
Db      572 CCGCGAGAAATCTACGCGCGGATGGAATCATCAAGCGCGTGTGCACTGAGCGGCTAC 631
Qy      1172 GTGAGCGGTGCAAAAACCTCAAGGTTATTAGCCGGTAGAGGCAAGCGATGTTTTT 1231
Db      632 GACCGATGCTGATCGTGGCTGTAGCGGTGCTCATGCAAGCTCCGGAACCGCTTCGCGAC 691
Qy      1232 CTCCAACAATCTCTGGGTGCA---CTTCCGGTATGAGGCGCATTTGCCGCTTCTGTG 1288
Db      692 CGCAGACGCCCGCCCGCTGACCCGATGTCACAGTGGACCTCGTGTGATCGCATCGCGCC 751
Qy      1289 ACCGCAAGGCTGGAATCGGGAGATTATGAAAGCTTTGGGGGACAGACTGTCAAGC 1348
Db      752 GCGCGAGCCGCGCGGCTCAGCGACTACTGCGACAGCGCGCCACCGAGCGGATGTTGT 811
Qy      1349 TTTCTGTGCGCGTGTGCGGCATTTAGCCCGGCTGTGCGGTTCCCTGCGCGCATGCG 1408
Db      812 TGTGAGGAACCGATCAAGAGGAAGCG---GGACCAAGCTGTGCGGCTGGCCGCTTTTG 868
Qy      1409 GGGCGCGGCGGTGTGTGATTTGGGAG-----GATCGCGCTTCAAGC 1453
Db      869 GCGCGCAACCCACGCTGTCTCTCGGCGAGCAAAAGGAGTAGAGCGGGGGAAGCACTG 928
Qy      1454 TTGGGCGGAGAGACTGCGTTTGGCGGTGCGTGAATTTGCGTGGCGGCGAGCTAAAC 1513
Db      929 TCGGCGCGCTGCGCTTACGAGAACCCGAGCGGAGTGGGCTTCCGCGGAGCTGTGCC 988
Qy      1514 TGCCGATGCTGCTCATCATGACACCTTCGCGCGGATTTGTCAGAGCGGCTGAGAGAC 1573
Db      989 TGCCGCTGTGTGTGATTAAGCGCGGCGGACCGGCTGTGCGCGGCGGACGCAACAG 1048
Qy      1574 TCGCATGCGAAGCTGATTTGCGGCACTTTTTCGAAGCTTATGACGCTTCCCTCCCA 1633
Db      1049 GCGGGCTGGCGCGGACAGATGCGCATTTGCTGCGCGAGCTGTCAACCTGGATCCCGGA 1108

```


Db 1007131 GCGGGCTGGCGGCGGAGATTCGGGATTCGCTGGCGGAGCTGTCTACGCTGGATACCCGGA 1007072
Qy 1634 CCGTTTCGCTATTATTTGTGTAAGGCGCTTGGCGGTGGCGCGCTGCATGCTGCCGCGG 1693
Db 1007071 CCGTGTTCATCTCTGTGGGCGGAGGCGGCGGCGCGCTGTGGCGATGTTGGCCGCGG 1007012
Qy 1694 ATTCGCTAGCGGCGGCGGAAACCGGCTGTGCTCCGCAATTCGACACGAGGCGCGCTCGG 1753
Db 1007011 ACCGGGTCTGGCGGCTCACTCCAGCGGCTGGCGCTGCTGCTTGCCTCCGAGGAGCGAGCG 1006952
Qy 1754 CCATCTCTTCCGCGGACCAACCAACGCGCGGAAATCATAGAGCGAGCGGCTGCAGG 1813
Db 1006951 CGATGCTGTTCGAGAGACTGCTCATGCGCGGAACTGCTGCGCGCCCAAGGCAATCCGCT 1006892
Qy 1814 CGCAGCGACTTTTAAGCCAAAGGCTTATTCGACGGGATGCTGCCGGAACCGAGCACTTTG 1873
Db 1006891 CGGCGGACCTACTAGTATGCGGGGATTTGACACCATGCTGCGGAGTACCCCGACGCGG 1006832
Qy 1874 TTGAAGA 1880
Db 1006831 CAGACGA 1006825

RESULT 8
AI199682/c
ID AI199682 standard; DNA; 4411529 BP.
XX AI199682;
AC AI199682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
PI Fleischmann RD, White OR, Fraser CW, Venter JC;
XX
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
PS Claim 3; SEQ ID NO 1; 3bp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AI199683) and
CC H37Rv (AI199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.

XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 14.7%; Score 312.6; DB 22; Length 4411529;
Best Local Similarity 54.0%; Pred. No. 1,4e-82;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
Qy 572 ACTGGTTTGGACCCCTGACAGCTTCACTTCTTGGAAATGAAATCCCAATATGACAAACC 631
Db 1008147 ACCGGGTGTAGACCCGGGATCTTCTTCTGACGTGGATACGAGCGCTGGCGGTCCGG 1008088
Qy 632 TCAATCAAGCTATGACAGACCTTGGAGCGGGCTGGAAGCAAGCCAAATGCGATTAAT 691
Db 1008087 TAGCCGACTCTTATGCGCGGAGCTGAGCCGCTGCGGCGGCGCACCAGCGGAGCAAT 1008028
Qy 692 CGGTAAATTAATGAGAAAGCAACCGTGAAGGCAATTCGATAGCCGTTATTTTTCGATTT 751
Db 1008027 CGGTGACAGACCGGTAAGGAGACGCGTATTCGGGCGGCGGCTGCTGTGTGAGT 1007968
Qy 752 TTTCTCTCTCGGCGGCTTCTTGGGACGCTCCGCTGCTGCGCATATGAAGCGCATTC 811
Db 1007967 TCGACTTCTGCGGCGGCTGATTTGGGATGACCGCGCAACGATACCGCGCGCTCG 1007908
Qy 812 ACCGCGCACAGAGCTGAATCTCCACTGCTGCTTCCCTGCTTCCGCTGCTGCGGCA 871
Db 1007907 AGCGGGCGACCGCGCGAGCGGCTGCGTACTGACGCTACCAAGCTCGGAGGCAACCGCA 1007848
Qy 872 TGCAGAAAGCAATCGAGCTTTTGTCAATGATGATCAATACCGCGCTGACGCTC 931
Db 1007847 TGCAGAAAGCAACGATGCGCTTTCTGAGATGATGAAATCGCTGCGGCGCATCGCTG 1007788
Qy 932 ACCGCGAGGCGCATTTGCGGCTCTGCTGATTTTGGCAATTCACAGATGGTGGCGCCA 991
Db 1007787 ACAACAGGCGCGCTGCTGCTTACTGCTGATTTTGGCCATTCGACCAAGGATGTTT 1007728
Qy 992 TGGCTTGTGGGCTTCACTGCGCATCTCACTTTTGGGAAACCGCGCGCAATAGTT 1051
Db 1007727 TCGGTGTGTGGGCTCGCTGGGCACTCACCGTCCGAGCGGCGGCTGATCGCT 1007668
Qy 1052 TCTGTGGCTCTCGCTGTGTGAGTTAACATGCGGATGCGCTTCCAGAGCTGTGAGC 1111
Db 1007667 TTTGTGGGACACCGGCTTATGATGCTTCTATGAGCAACCCCTTCCATCGGCGTCCAA 1007608
Qy 1112 AGCGGAGATTTTGTGAAACCTGTGTGATGATGAAATGTGTGCCACTCCATTTG 1171
Db 1007607 CCGCGGAAATCTACCGCGGCAATGATCATGACGCGCTGTTGCACTGACCGGCTAC 1007548
Qy 1172 GTGAGCGGTGGCAAAACCTCAAGTTATTCAGCCGCTGAGGCAACGATCTGTTT 1231
Db 1007547 GACCGATGCTGATTCGTGCTTACGCTGTCTATGACGCTCCGAAACGCTTCCGGCAC 1007488
Qy 1232 CTCCAACTCTCTGGGCTGCA---CTTCCGCTGATGAGAGGATTTGCGCTTCTGTG 1288
Db 1007487 CGGAGACGCGCGCGCGCTTACCGGATGCTCCAGTGGGCTGTGTGTGCAATCCGCGC 1007428
Qy 1289 ACCGCGAGAGCTTGAATGCGGAGATTTATGAAACGTTGGGGGCAAGCTGTCAAGC 1348
Db 1007427 GGCAGACGCGCGGCGGTAGGACGCTACTGCAACGCGCGCACGACGCGGCTGTGT 1007368
Qy 1349 TTTCTGTGCGCTGTGCTGCGATTGAGCCCGGCTGTGCGCTTGTGCGCGCATCG 1408
Db 1007367 TGTGAGAAACGATCAAGGGAAGCG---GCAACACGCTGCTGGCGCTGCGCGCTTGG 1007311
Qy 1409 GGGGCGGCGCGGTGTGTGATTTGGGCGA-----GATCGCGCTTACGCG 1453
Db 1007310 GCGGCGCAACCGAGGTGTCTCTGCGCAGCAAGGCGAGTAGCGCGGCGGGAAGCATG 1007251
Qy 1454 TTGGGCGGAGAGCTGCGTTTGGCGGTGCGGATTTGCGTGGGCGCGGAGCTAAAC 1513
Db 1007250 TCGGGCGCGCTGCTTACGGAAGCCGAGCGGAGTGGGCTGCGCGCGAGCTGTGCC 1007191
Qy 1514 TGGCGATGCTGCTCATCATGACACTCGCGCGCAATTTGTGCGAGCGGCTGAGAGC 1573

Db 1007190 TCCCGCTGCTGTCATTCAGCCGCGACCCGCTTGTGCGCCGACGCGAAGCAG 1007131
QY 1574 TCGGATCGCAAGCTCGATTGGCGGACCTTGTCCAAAGCTTATTCAGCGTCCCGCCCA 1633
Db 1007130 GGGGCTGCGCGGCGGATCGGCGATTGCTGGCCGAGCTCGTACGCTGATACCCCGA 1007071
QY 1634 CCGTTTCGTCATTATTTGTCAGGGCGTGTGGCGCGCTGGCGCATGCTGCCCGCG 1693
Db 1007070 CCGTCTCATCTGCTGGCGCGAGGCGAGGCGCGCGCGCTGGGATGTTGCCCGCG 1007011
QY 1694 ATCTGCTACGCGGCGGAAACGCGTGTCTCCGATTCGCACCAAGGCGCGCTCG 1753
Db 1007010 ACCGGGTGTGCGCGGATCTCCAGCGCTGGCGCTTGTGCTCCGAAAGGAGCGAG 1006951
QY 1754 CCATCTCTTTCGCGGACCAACCAAGCGCGGAAATCATAGAGCGAAGCGCTGCAAG 1813
Db 1006950 CGATCTGTTTCCGAAGCACTGCTCATGCGCGCACTCGCTGCCCGCAAGGCATCCGT 1006891
QY 1814 CGCAGCACTTTAAGCCAGGCTTATTCAGCGGATCTGCGCGAACCAGCACTTTG 1873
Db 1006890 CGGCGACCTACTGAAGTGGGATTTGTGACACCATCTGCGGAGTACCCCGAGCGCG 1006831
QY 1874 TTGAAGA 1880
Db 1006830 CAGACGA 1006824

RESULT 9

ACA00248
ID ACA00248 standard; DNA; 168 BP.

ACA00248;

02-JUN-2003 (first entry)

C. glutamicum derived ORF SEQ ID 239.

Coryneform; nucleic acid array; fermentation; culture; ds.

Corynebacterium glutamicum.

DE10128510-A1.

19-DEC-2002.

13-JUN-2001; 2001DE-1028510.

13-JUN-2001; 2001DE-1028510.

(DEGS) DEGUSA AG.

Farwick M, Moekel B, Pfeifferle W, Bache B, Hutmacher K;

WPI; 2003-279970/28.

New nucleic acid array useful for monitoring mRNA expression of
Corynebacterium glutamicum during fermentation, comprising nucleic acid
from Corynebacterium glutamicum -

Claim 1; Page 154; 709pp; German.

This invention describes a novel nucleic acid array involving
Corynebacterium glutamicum polynucleotides. The arrays are used to
analyse C. glutamicum, particularly for monitoring a fermentation process
to determine expression levels of C. glutamicum cellular mRNA. Such
monitoring particularly differentiates between expression levels of
different strains of C. glutamicum and allows the adjustment of different
culture and fermentation conditions. ACA0010-ACA02188 represent
C. glutamicum derived polynucleotides described in the disclosure of the
invention.

Sequence 168 BP; 51 A; 40 C; 28 G; 49 T; 0 other;

Query Match 7.9%; Score 168; DB 25; Length 168;
Best Local Similarity 100.0%; Pred. No. 6.3e-41;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ATGTTTTCACCAATCTGGCGGACGACCAAGATATGTTGTGATCAGACTAAAGCT 125
Db 1 ATGTTTTCACCAATCTGGCGGACGACCAAGATATGTTGTGATCAGACTAAAGCT 60
QY 126 GTGAATGTGAAGTTACCTTAACATTCATTCGATGCGATAGCGATTTGGAAAACTACTCC 185
Db 61 GTGAATGTGAAGTTACCTTAACATTCATTCGATGCGATAGCGATTTGGAAAACTACTCC 120
QY 186 CCCCAATATCTTAACCTTAAGTATGTTTACCTGATTTA 233
Db 121 CCCCAATATCTTAACCTTAAGTATGTTTACCTGATTTA 168

RESULT 10

AAH68407/C
ID AAH68407 standard; DNA; 354 BP.

AAH68407;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 3442.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis; ds.

Corynebacterium glutamicum.

EP108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOGYO KK.

WPI; 2001-376931/40.

P-PsDB; AAG93188.

Novel polynucleotides derived from Coryneform bacteria, for identifying
mutation point of a gene, measuring expression of a gene, analysing
expression profile or pattern of a gene and identifying homologous gene

Claim 1; SEQ ID NO: 3442; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein
sequences from the Coryneform bacterium Corynebacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of coryneform bacterium, measuring expression amount and
analysing the expression profile or expression pattern of a gene derived
from Coryneform bacterium, and identifying a homologue of a gene derived
from coryneform bacterium. Coryneform bacteria are useful for producing
amino acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present sequence is a nucleic acid described
in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
European Patent Office.

Sequence 354 BP; 77 A; 97 C; 106 G; 74 T; 0 other;

PI Fontana MR, Pizze M, Masignani V, Monaci E;
 XX
 DR WPI: 2003-058415/05.
 DR P-PSDB; ABP78030.
 XX
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
 PT medicament for treating or preventing *N. gonorrhoeae* infection -
 XX
 XX Disclosure: Page 377; 815pp; English.
 XX
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records AB237706-AB242016 represent nucleic acid
 CC molecules of the invention.
 XX
 XX Sequence 876 BP, 183 A, 216 C, 272 G, 205 T, 0 other;

Query Match	5.8%	Score 122.2;	DB 25;	Length 876;
Best Local Similarity	51.6%;	Pred. No. 1.9e-26;		
Matches 280;	Conservative 0;	Mismatches 263;	Indels 0;	Gaps 0

QY	628	AACTCATCAAGGCTATGACAGAGACCTTTGAGCCGGCTCGAAGCAAGGCTCAATTCGAT	687
Db	274	AAAGACGCAAAAAAATATCCGAGCCGTTGATGCGCGACGCAAGCTGACCCGGAGAGAT	333
QY	688	GAATTCGGTAATTACTGAGAAAGGACCCGTGAGGGCAATTCGGTAACCGTTAATTTGTCC	747
Db	334	GACCGCGCTGGTGATGATGAAGGGAGTGAATGAACGGGTGTCGCCGTGTCGGCGGTTT	393
QY	748	GATTTTCTCTCTCGGCGGTTCTTTGGGCACGGTCCGTCCGTGCGATCATGAAGGCG	807
Db	394	GAATTCGCGTTATCGGCGGTTTCGATGAGGTTCGGTTGGGCGAAGCTTTCTGTACAGGGT	453
QY	808	ATTACCGCGCCACAGAGCTGAACTCCACACTGCTGATCCCTCGCTTCGGATGATGGC	867
Db	454	GTCGCGCGTGGGCTTGGCCGATTAATGTCCGTTTGTCTGTGTGGCGGCTTCCGGCGCGCG	513
QY	868	CGCATGCAAGAAACAATCGACTTTTGTATGATGATGATGATCCATAACCGCGGCTGCA	927
Db	514	CGTATGCAAGAGGGGTGTAACCTCGCTGATGACAGATGAGAGAAACAAGTGCCTCGCTCAT	573
QY	928	CGTACCGCGAGCGCATTTTGCCGTTCTCTGCTGTAATTTGCGCAATCCACGATGGGATGC	987
Db	574	TTGCTGACGGAAGAAACGCTGCCGTTTATTCGAGTGTGACCGCATTCGACTGATGGCGGCG	633
QY	988	GCCATGGCTCGTGGGGGTTTCATCTCGGGGATCTCATCTTTTGGCGAACCCGGGCGCGAGATA	1044
Db	634	GTAATCTCGACAGCTTCGCAATTTTAAAGCGCATGTCTGCTTGCAGAACGAAACCGCTGTATC	693
QY	1048	GGTTTCTCGAGTCCCTCGCGTGGTGAATTAACAATCGAGGACATGCGCTTCCAGACGGTGG	1107
Db	694	GGTTTTCGCGGTCCGCGCGGTGATTTGACAGACGGTCCGGAACGCTGCGGAGAGGCTTC	753
QY	1108	CAGCAGCGGAGAAATTTGGTAAAACTGGTGTGATTTGATGCAATTTGTGTCCCACTCCAA	1167
Db	754	CAACGCGCCGAGTTCTTGCTGAAAAAAGGCGGATTCGACAGATTTGTGACCGCGCGAT	813
QY	1168	TTTG	1170
Db	814	ATG	816

RESULT13	
AAS53191	
XX	AAS53191 standard; DNA; 867 BP
AC	
XX	AAS53191;
DT	13-FEB-2002 (first entry)

XX Enterococcus faecalis DNA for cellular proliferation protein #619.
DE
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
KM antibiotic; antibacterial; drug design.
XX
XX
XX Enterococcus faecalis.
OS
XX
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haseelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
XX P-PSDB; AAU53332.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
XX Claim 27; Seq ID No 6828; 51pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 867 BP; 266 A; 145 C; 209 G; 247 T; 0 other;
XQ

Query Match	5.7%;	Score 121;	DB 23;	Length 867;
Best Local Similarity	53.0%;	Pred. No. 4.3e-26;		
Matches 259;	Conservative	0;	Mismatches 230;	Indels 0;
			Gaps	0

Qy	688	GATGAATCGGTAAATTACGTGAGAAAGGACCGGTGAGGGCAATTCGCGTACCGTATTATTTG	744
Db	328	GATGAAGCGTGTCTTAACGGGAGAAAGCAACATTGAAGACACAGTGTGTGCAATTGGAAATC	387
Qy	745	TCCGATTTTCTCTCTCTCGCGCGGTCTTTTGGGACACGTCGCGTCGTCGCAATCAAGAA	804
Db	388	ATGACGCTAATTTTATATATGAGGACGATGGGAAACGATTTGTGTGAAAAATCAACCC	447
Qy	805	GGGATTCACCGCGCACAGAGCTAATACTCCCACTGCTGTCTCCCTGCTTCGCGGT	864
Db	448	TTGTATTGACGGGACACAGAAACATTATACCATATGATTTTCACTGATCTCGTGTGT	507

QY 865 GCGGCGATGCGAAGCAATCGAGCTTTTGTATGATGCTGTCCTTAACCGGCGCTGTG 924
 Db 508 GCCCGTATGCAAGAAAGAAATTTTTCATTGATCAAAATGCGAAAATTTGCGCGCTTTG 567
 QY 925 CAGCGTCACCGCGAGGCGCATTTTGCCTTCCTGTATTTGGCGCAATCCGAGATGGT 984
 Db 568 CAACGGCATTAACAAGAGAGCTTGCTGTATCTTACGGTATGACTGATCCAAACGACTGGC 627
 QY 985 GGGGCGATGCGCTTCGTGGGTTTCATCTGGGATCTCACTTTTGGGAAACCGCGCGCAG 1044
 Db 628 GGGTTACCGCAAGTTTGGATGATGCGATATATTTTGGCAAGCCTCAGAGTTTA 687
 QY 1045 ATAGTTTCTCTGGGCTCTGCGCGTGTGAGTTAACCACTGGGATGCGCTTCCAGACGT 1104
 Db 688 ATCGTTTGTCTGCGCGCGCTGTATTAATGAACAACATTCCTCAAGAGTTGCCAGATGAT 747
 QY 1105 GTGCGACGAGCGGAGAAATTTGTGTAACACTGTGTGATGATGAATTTGTGCGCATCTC 1164
 Db 748 TTTCAAAAAGCGAGTTCTTTTGAACATGTTTGTGATCAGATTGTGCGAAGAAAT 807
 QY 1165 CAATTGCGT 1173
 Db 808 CTTTTCGCT 816

RESULT 14

AA13395

ID AA13395 standard; DNA; 6021 BP.

AC AA13395;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:458.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;

KN vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN WO980555-A2.

XX 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.

PS Claim 1; Page 1691-1694; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AA13395 to AA13399 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.

SQ Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;

Query Match 5.7%; Score 121; DB 20; Length 6021;

Best Local Similarity 53.0%; Pred. No. 1,4e-25;

Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 685 GATGAATCGGTAATTAATCTGGAAGAGCAACCTGAGGAGCAATCCGATGCGTTATTTTG 744
 Db 5321 GATGAACCTGCTTAACGGGAGAAAGCAACATTAAGAGCAACCTGATGCAATGGAATC 5380
 QY 745 TCCGATTTTCTCTCTCGCGGCTTCTTGGGACGCTGCGTGGTGGCCATCAAGAG 804
 Db 5381 ATGAGCGCTAATTTTATCATGAGGAGTATGGAACGATTTGTGTAAGAAAAATCACAGC 5440
 QY 805 GCGATTACCGCGCAAGAGCTGAAATCCACTGCTGCTCTCCCTGCTTCCGTTG 864
 Db 5441 TTGTTGAGCGGCGGAGCAAGAAAGCAATTAACAGTATGATTTTCACTGATCTGCTG 5500
 QY 865 GCGGCGATGCGAAGAAACAAATCGAGCTTTGTATGATGCTGTCCTTAACCGGCGCTGTG 924
 Db 5501 GCCCGTATGCAAGAAAGAAATTTTTCATTGATGCAAAATGCGAAATTTGCGCGCTTTG 5560
 QY 925 CAGCGTCACCGCGAGGCGCATTTGCCGTTCTGTGTATTTGGCGCAATCCAGATGGGT 984
 Db 5561 CAACGGCATTAACAAGAGAGCTTGCTGTATCTTACGGTATGACTGATCAACAGACTGGC 5620
 QY 985 GGGGCGATGCGCTTCGTGGGTTTCATCTGGGATCTCACTTTTGGGAAACCGCGCGCAG 1044
 Db 5621 GGTGTTACCGCAAGTTTGGATGATGCGCATATATTTTGGCAGAGCTCAGAGTTTA 5680

QY 1045 ATAGTTTCTCTGGGCTCTGCGCGTGTGAGTTAACCACTGGGATGCGCTTCCAGACGT 1104
 Db 5681 ATCGTTTGTCTGCGCGCGCTGTATTAATGAACAACATTCCTCAAGAGTTGCCAGATGAT 5740
 QY 1105 GTGCGACGAGCGGAGAAATTTGTGTAACACTGTGTGATGATGAATTTGTGCGCATCTC 1164
 Db 5741 TTTCAAAAAGCGAGTTTCTTTTGAACATGTTTGTGATCAGATTGTGCGAAGAAAT 5800
 QY 1165 CAATTGCGT 1173
 Db 5801 CTTTTCGCT 5809

RESULT 15

ABS99190

ID ABS99190 standard; DNA; 6021 BP.

AC ABS99190;

DT 18-DEC-2002 (first entry)

DE Enterococcus faecalis contig sequence #458.

XX Computer readable medium; Enterococcus faecalis; microbe; growth;
 KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KW therapeutic; industrial; fermenting; sugar source; metabolic; vaccine;
 KW biotech technology; antibacterial; modulator of nucleic acid expression;
 KW contig; ds.

OS Enterococcus faecalis.

PN US2002120116-A1.

XX 29-AUG-2002.

PF 04-MAY-1998; 98US-0070927.

PR 04-MAY-1998; 98US-0070927.

XX (KUNS/) KUNSCH C A.
PA (DILL/) DILLON P J.
PA (BARA/) BARASH S.
XX

PI Kunsch CA, Dillon PJ, Barash S;
XX

DR WPI, 2002-750065/81.
XX

XX Computer readable medium having recorded on it a Enterococcus faecalis
PT nucleotide sequence useful for detecting diseases related to
PT Enterococcus infections in animals
XX

PS Claim 1; Page -; 119pp; English.
XX

CC The present invention relates to a new computer readable medium with an
CC Enterococcus faecalis nucleotide sequence. The invention is useful to
CC diagnose the presence of E. faecalis in a sample or determining the
CC presence of a specific microbe in a sample. The invention is also useful
CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine,
CC to confer resistance to Enterococcal infection, for commercial,
CC therapeutic and industrial purposes, and for fermenting a particular
CC sugar source or to produce a particular metabolite. The invention is
CC useful for detecting diseases related to Enterococcus infections in
CC animals, and for detecting E. faecalis using biotech technology. The
CC present nucleic acid sequence represents an Enterococcus faecalis con-
CC DNA sequence of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification but was obtained in electronic format directly
CC from USPTO at <http://seqdata.uspto.gov>.
XX

Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;

Query Match 5.7%; Score 121; DB 24; Length 6021;
Best Local Similarity 53.0%; Pred. No. 1.4e-25;
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 685 GATGATGCGTAACTACTGAGAGAGGACCGTGAGGCGCATCCGGTAGCCGTTATTTTG 744
DB 5321 GATGAGCTGTCTTAAAGGAGAGACAAATTTAAAGCAAGCTGTGCAATGGAATC 5380
QY 745 TCCGATTTTCTCTCTCGCGGTTCTTTGGACGCGTCCGTCGCGCATGAAAG 804
DB 5381 ATGACGCTAATTATATCATGGCAGTATGGACGATTTGTGTAATAAATCACAGC 5440
QY 805 GCGATTACCGCGGACAGAGCTGAATCCCACTGCTGCTCCCTGCTTCCGCTGT 864
DB 5441 TTGTTGAGCGGCGACAGAAAGCATTTACAGTAGTATTTTCACTGCACTGCTGTGT 5500
QY 865 GCGGCGATGAGAGAGACATGAGCTTTTGCATGATGCTGTCATTAACCGGCGCTGTG 924
DB 5501 GCGCGTATGACAAAGAGAAATTTTTCATTTGATGCAAAATGCGAAATTTGGCGCTTTG 5560
QY 925 CAGCGTCAACCGGAGCGGCGATTTGCGGTTCTGCTGTAATTTGCGCAATCCACGATGGGT 984
DB 5561 CAACGCGATTAACAAGCAGGCTTCTGCTATCTTAAGGTAATGATGATCAACGACTGGC 5620
QY 985 GCGGCGATGCGCTCTGTTGGGTTCACTGCGCATCTCACTTTTTCGGAACCGGCGCGCAG 1044
DB 5621 GGTGTTACCGCAAGTTTTCGATGATGCGCATATTTTGGCAGAGCCTCAGAGTTTA 5680
QY 1045 ATAGGTTTCTGCGCTCGCGGTGATGAGTAAACACATGCGGCGATGCGCTTTCAGACGCT 1104
DB 5681 ATCGGTTTCTGCGCGCGCGCTGTAATTTGAACAACGATTCGTAAGAGTTGCGCAGATGAT 5740
QY 1105 GTGCGACAGCGGAGAAATTTGTGAAACTGCTGATGATGATGATTTGTGCGCACTC 1164
DB 5741 TTTCAAAAGCGCGAGTTCTTTTAGAACAATGTTTGTAGATCAGATTGTGCCAAGAAT 5800
QY 1165 CAATTGCGT 1173
DB 5801 CTTTTCGCT 5809

Search completed: November 12, 2003, 20:25:38
Job time : 625.017 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using gw model

Run on: November 12, 2003, 18:57:58 ; Search time 4673.43 Seconds
(without alignments)
11040.790 Million cell updates/sec

Title: US-10-024-370-1

Perfect score: 2123
Sequence: 1 ctccgagcgggagtcggtgat.....gacctcatgctcggaattc 2123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estpl:*
7: em_estpro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	4.1	488	28	BH379574 AG-ND-133
2	86.2	4.1	707	13	B0654446 111213FO
3	84.2	4.0	604	9	AW584783 N210917e
4	82.6	3.9	773	12	BI311746 EST531349

5	81.2	3.8	1063	29	B2548722	B2548722	pac91-60
6	81	3.8	712	12	BI310727	BI310727	EST531247
7	79.8	3.8	580	14	CA990607	CA990607	EST644115
8	79.8	3.8	792	10	BG584921	BG584921	EST486683
9	78.8	3.7	1042	29	B2561247	B2561247	pac92-164
10	78.4	3.7	539	12	BI725661	BI725661	1031080C0
11	77.6	3.7	652	29	AG247247	AG247247	locus_jap
12	75	3.5	617	14	CB977199	CB977199	CAB4003
13	74	3.5	668	9	AI229631	AI229631	EST26326
14	73.8	3.5	677	28	B2035752	B2035752	oen49c05
15	73.6	3.5	606	14	CB346970	CB346970	CAB25G00
16	73.4	3.5	453	14	CB081454	CB081454	hk51h06.9
17	73.4	3.5	708	28	B2051213	B2051213	jnr62h07
18	72.2	3.4	675	28	BH988285	BH988285	oe167b06
19	72.2	3.4	696	28	B201471	B201471	oe167b06
20	72.2	3.4	700	28	B2033778	B2033778	oe101d09
21	72.2	3.4	703	28	B2017314	B2017314	oe167c12
22	72.2	3.4	704	28	BH930860	BH930860	od101b11
23	72.2	3.4	709	28	B2075724	B2075724	1k153c01
24	72.2	3.4	711	28	AQ957569	AQ957569	LERQA06TF
25	72.2	3.4	752	28	B2033142	B2033142	oen45a09
26	72.2	3.4	781	28	B2071328	B2071328	1j113h07
27	72.2	3.4	807	28	BH715901	BH715901	BOHYJ64TR
28	72.2	3.4	844	28	BH462655	BH462655	BOGR54TF
29	71.8	3.4	1872	28	BH770899	BH770899	LMGCa963
30	71.4	3.4	607	14	CB977121	CB977121	CAB40003
31	71.4	3.4	637	28	AQ964265	AQ964265	LERGU88TF
32	70.4	3.3	843	28	BH477984	BH477984	BOHQP36TF
33	70	3.3	719	28	B2052334	B2052334	jnr6b12
34	69.4	3.3	697	28	B2006948	B2006948	oen84f09
35	69	3.3	405	28	BH397912	BH397912	AG-ND-143
36	68.8	3.2	627	9	AU293522	AU293522	AU293522
37	68.4	3.2	692	28	B2026810	B2026810	oen127c06
38	68.4	3.2	748	28	BH987475	BH987475	oej71e06
39	68.2	3.2	523	28	AQ991181	AQ991181	RfC02065
40	68	3.2	698	28	BH994577	BH994577	oe162f03
41	67.8	3.2	452	9	AW587803	AW587803	ST67B1.P
42	67.4	3.2	722	28	BH925341	BH925341	od146f02
43	67	3.1	886	29	B2553868	B2553868	pac91-60
44	66	3.1	1333	29	B2579621	B2579621	meh2.686
45	65	3.1	558	12	BI725648	BI725648	1031080B1

ALIGNMENTS

RESULT 1
BH379574
LOCUS
DEFINITION
AG-ND-133C23.TF ND-TAM Anopheles gambiae genomic clone AG-ND-133C23
' genomic survey sequence.
ACCESSION
BH379574
VERSION
BH379574.1 GI:17325716
KEYWORDS
GSS.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 488)
Hong Y.S., Hogan J.R., Wang X., Sarkar A., Sim C., Loftus B.J., Ren C., Huff E.R., Carlile J.L., Black K., Zhang H.-B., Gardner M.J. and Collins F.H.
Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
Mol. Genet. Genomics 268 (6), 720-728 (2003)
JOURNAL
MEDIUM
PUBMED
12653598
COMMENT
Other GSSs: AG-ND-133C23.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@ctr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seg primer: M13 For

FEATURES

source

Location/Qualifiers
1..488
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-133C23"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 137 a 86 c 109 g 156 t

ORIGIN

Query Match 4.1%; Score 88; DB 28; Length 488;
Best Local Similarity 49.0%; Pred. No. 3,4e-14;
Matches 235; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY AGGCTATGAGACCTTGGAGCGGCTCGAAGGCAATCGATGATCGTAT 698
DB 1 ACGTTATCTGACCCATTAAAGAGTAAAGGCAAGCTAAATTACGATTCGTACG 60
QY TACTGAGAAGGACCGTGGAGGCAATCCGCTAGCCGTTATTTGTCGATTTTCCT 758
DB 61 TATCTGTTGGTAAAGTAAAGGACGATGACCTGTTGTTTCATGATGATTCCTTT 120
QY CCTGCGCGGTTCTTTGGGACGCGTCCGCTGCGTGCATGAAAGGCAATCCGCGC 818
DB 121 TATCGAAGGCTATTATGCTGTGTAAGGAGAAAAAATCATCCGCTATGCACTATGC 180
QY CACAGAGCTGAAGTCCCACTGCTGCTCCCTGCTCCGCTGCGGTCGCGATCAGGA 878
DB 181 AATGAAGCTAAATTACCTTTATGATTTATTCAGATCGTGTGAGCAAGATGACAGA 240
QY AGACAATCGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 938
DB 241 AGCAGCTTATTCATTAATGACAGTTGGCAAAAGTTCAAGGTAAGTACGCTTTCGGA 300
QY GGGCGATTTGCGGCTTCTGCTGATTTGGGCAATCCGACAGATGGTGGCGGCGCTC 998
DB 301 CAAGGCTTACCATATGTTGCTTATTTAACTGACCTTCTTTTGTGATTTACAGCTTC 360
QY GTGGGCTCATCTGGGCAATCTCACTTTTGGGACCCGCGCGCGATGATGATTTCTGG 1058
DB 361 TTTTGGATGATCTGCTGATATCATTTATGCTTAACCTGTGCTGATTTATGCTTTCAG 420
QY 1059 TCCTGCGGCTGATGATTAACCACTGGGATCGCTTCGACGCGTGTGACAGAGCGGA 1118
DB 421 TCCTGAGATTTCAAGAAACAATGATGATGATGATGATGATGATGATGATGATGAT 480

RESULT 2
BU654446 707 bp mRNA linear EST 30-SEP-2002
LOCUS 112131P06 y1 C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete
DEFINITION (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA
ACCESSION BU654446
VERSION BU654446.1 GI:23366627
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 707)

AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrago, J., Sillflow, C. and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Vascular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1112

JOURNAL Unpublished
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES

source

Location/Qualifiers
1..707
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="21gr (CC-1690 wild type mt+) & 6145C (CC-1691 wild type mt-)"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Gamete library was constructed by Hu Zhao, Jeffrey McDermott, William J. Sneli and John Davies. Strain 21gr cells (CC-1690, mating type plus) and strain 6145C cells (CC-1691, mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Sager and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period. PolyA mRNA was purified from each sample every 2 hours for the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 127 a 241 c 227 g 112 t

ORIGIN

Query Match 4.1%; Score 86.2; DB 13; Length 707;
Best Local Similarity 50.6%; Pred. No. 1,4e-13;
Matches 208; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 1449 CACGCTGGGCGGAGAGCTGCTTTGGCGCTGCGATTTCCGTGCGCGGAGCT 1508
DB 286 CATGCCCGGCGCAAGCGCTACCGCAAGGCGCTGCGTTATGCGCCAGCGCAAGTT 345
QY 1509 AAACCTGCGATCGTGTGATCATGACACCTCCGCGCGCAATTTGCGAGGCGCTGA 1568
DB 346 TGTCTGCGCATCATCTTCTGAGACAGCCCGAGGCTATGCGCGGCAAGACCGGGA 405
QY 1569 GAGCTGGCATCGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1628
DB 406 GAGCTGGCGGAGGCGGAGGCGATTTGCGTGAACCTGCGTGAAGATGTTCCGCTGCGGT 465
QY 1629 CCCACGCTTTGGTCAATATTTGCTCAGGCGGTTGGCGGCGCGCTGCGCATGCTGCC 1688
DB 466 GCCCATATCTCGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 525
QY 1689 GCGCGATGCTGATACCGCGCGGAAAGCGGTGCTGCGCATTTGCGACACGAGGCGCG 1748
DB 526 GCGCAACGCAACTGATCATGAGAAAGCGGTGTTACTAGTGGCTTCCGCGGCGGCTG 585
QY 1749 CTGCGCATCTCTTTCGCGGACCAACCAAGCGCGGCAATCATAGGAGCAAGGCGT 1808

Db 586 CGCGCCATCTGTGAGAGCCGCTCTGCCGCCGAGGCCACTGAGCCCTGGCAT 645
 Qy 1809 GCAGGCGCAGCACTTTAAGCAGGCTTATCGAGGATGTCGCCGA 1859
 Db 646 CACCTGCGCGAGCTGTGAAGTTCGCGCTCATGACCAATCTGCCGA 696

RESULT 3
 AM584783
 LOCUS
 DEFINITION N210917e MHAM Medicago truncatula/Gloms versifforme mixed EST
 library cDNA clone MHAM-7110, mRNA sequence.
 ACCESSION AM584783
 VERSION AM584783.1 GI:7261837
 KEYWORDS
 SOURCE Medicago truncatula/Gloms versifforme mixed EST library
 ORGANISM Medicago truncatula/Gloms versifforme mixed EST library
 Eukaryota; mixed EST libraries.
 REFERENCE 1 (bases 1 to 604)
 Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
 ESTs from roots of Medicago truncatula after colonization with
 Gloms versifforme
 TITLE Unpublished
 JOURNAL
 COMMENT Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Other name: MHAM-7b-805; Date: 3/14/00; Updated to the Database of
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is
 available at 'http://chrysis.camu.edu/medicago'.
 Seq primer: 73.

FEATURES
 source
 1..604
 Location/Qualifiers
 /organism="Medicago truncatula/Gloms versifforme mixed EST
 library"
 /mol_type="mRNA"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="MHAM-7110"
 /tissue_type="roots colonized with Gloms versifforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Gloms versifforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XL0LR"
 /clone_lib="MHAM"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Gloms versifforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 StrataGene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."

BASE COUNT 179 a 95 c 136 g 194 t
 ORIGIN

Query Match 4.0%; Score 84.2; DB 9; Length 604;
 Best Local Similarity 48.8%; Pred. No. 5e-13; Mismatches 268; Indels 3; Gaps 1;
 Matches 258; Conservative 0;

Qy 634 AATCAAGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCATGATCG 693
 Db 13 AATGATGATTAACAATGCTGCTTATCTTATCAAGACAGAACCGATCTGATCG 72
 Qy 634 GTAATTACTGGAAGACCACTGAGAGGCGATTCGGTATCCGTTATTTTGTCCGATTTT 753
 Db 73 GTTCAACAGGACAGCTCAAGTAATGATTTCTGTAGCAATAGTATATGATTTT 132

Qy 754 TCCCTCTCGCGGCTTTCTTGGGACAGGTCGCGTGGCCGATCATGAAAGCGATTCC 813
 Db 133 GAGTTATGGAGAGTAGATGAGATCCGTAGGAGTGAATAATCTCGTTGATTGA 192
 Qy 814 CGCGCCACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCCGAGTGGCGCATG 873
 Db 193 TATGCTACCAATCAACGTTTACCTCTTATTTATAGATGTCGTGAGAGACGGGTATG 252
 Qy 874 CAGAAAGCAATCGAGCTTTTGTCTCATGATGTCATTAACCGGCGCTGACGCTCAC 933
 Db 253 CAAGAAGAGAAAGTTGAGCTTATGCAAAATGGCTAAATTTCTTATATATATAT 312
 Qy 934 CGCAGAGCGCATTTCCGCTCTGCTGATTTTGGCA---TCCACAGATGGATGGCGCC 990
 Db 313 CAATCATCAAAAGTTATTTCTATATACGATCTTACATCTCTACTACTGAGGGTA 372
 Qy 991 ATGGCTGTGGGGTTCACTCGGCGATCTCACTTTTGGGAAACCGGCGCGCATAGT 1050
 Db 373 ACAGCTAGTTTGAAGTGTGGGGATATCATTTCTTCCGAAACCGATGCTTACATGCA 432
 Qy 1051 TTCTCGGCTCTCGCGTGTGAGATTACCACTGGCATGCGCTTCCAGACGCTGCA 1110
 Db 433 TTTGCAAGTAAAGATTAATGAAAGAACTTGAAGATTGAAGTCCCGAAGTATACAA 492
 Qy 1111 CAGCGGAGAAATTTGGTAAACTGCTGATGATGATGATTTGTGCG 1159
 Db 493 TCGGCTGAATTTTATTCGAAAGGGTGCATTTGATCTGATCCAC 541

RESULT 4
 B1311746
 LOCUS
 DEFINITION EST5313496 GSD Medicago truncatula cDNA clone pGSD15E20 5' end,
 mRNA sequence.
 ACCESSION B1311746
 VERSION B1311746.1 GI:14986073
 KEYWORDS
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula (barrel medic)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 773)
 Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
 J., and Fraser,C.M.
 ESTs from developing reproductive tissues of Medicago truncatula
 Unpublished
 JOURNAL
 COMMENT Contact: Michael A. Grusak
 USDA/ARS Children's Nutrition Research Center
 Baylor College of Medicine
 1100 Bates Street, Houston, TX 77030-2600, USA
 Tel: 713-798-7044
 Fax: 713-798-7078
 Email: mgrusak@bcm.tmc.edu
 B398601e
 TIGR sequence name: MTPA234TK
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA GTG GAT CC).

FEATURES
 source
 1..773
 Location/Qualifiers
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="pGSD15E20"
 /tissue_type="immature seeds"
 /dev_stage="Immature seeds, 11 to 19 days after
 pollination"
 /clone_lib="GSD"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Immature seeds, collected from pods ranging in age
 from 11 to 19 days after pollination, were harvested from

greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the UniLap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

BASE COUNT 261 a 112 c 154 g 246 t
ORIGIN

Query Match 3.9%; Score 82.6; DB 12; Length 773;
Best Local Similarity 48.6%; Pred. No. 1.7e-12;
Matches 257; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

```

Qy 634 AATCAAGGCTATGCAGAGACCTTGGAGCGGCTCGAACCAAGGCCAATGGGATGATCG 693
Db 33 AATGATGATTTCAAAATCGTCTTATTCTTATCAAGACAGAACCGATTACTGATGCG 92
Qy 694 GTAATTACTGAGAAAGCAGCGTGAAGGCATTCGGTAGCCGTTATTGTCGATTTT 753
Db 93 GTTCAACAGGCGACAGTCAAGTAATGTAATCTCTGAGCAATGATGATTATGATTTT 152
Qy 754 TCCCTCCTGGCGGCTCTTTGGGACAGGCGCGTGGCGCATGATGAAGCGCATTCAC 813
Db 153 GAGTTTATGGAGAGTATGATGAGATCCGATGAGGTGAGAAATACCTCGGTTGATGAA 212
Qy 814 CGCGGCACAGAGCTGAACTCCACTGCTGCTCCCTCGCTCCGATGAGCGGCATG 873
Db 213 TATGCTACCAATCAACGTTTACCTTATTAATGATGATGCGTGGAGAGCGGATG 272
Qy 874 CAGAGACATTCAGCCTTTTGTATGATGATGATGATGATGATGATGATGATGATG 933
Db 273 CAGAGAGAGGTTGAGCTTAATGCAAAATGCTAAATTTCTCTTTTATTAATATAT 332
Qy 934 CGCGAGCGCATTTGCCCTCTCTGATTTTGGCAAA---TCCACAGATGGGTGGCGCC 990
Db 333 CAATCATCAAAATGTTATTTATGATGATGATGATGATGATGATGATGATGATGATG 392
Qy 991 ATGCGCTCGGAGGTTCACTGAGCATCTCACTTTGGGAAACCGGCGCGCATAGAT 1050
Db 393 ACAGCTAGTTTGGAAATGTTGGGGATATCAATTTCTTCCGAACCGGATGATCAACA 452
Qy 1051 TTCCTGGGCTCTCGCGTGGTGAATTAACACTGGCGCATGGCTTCGACAGGTGTGAG 1110
Db 453 TTGCGAGTTAAAGATGATGAGAAACGTTGAAGATCGAAGTCCCGAAGGATACAA 512
Qy 1111 CAGGCGGAGAAATTTGGTGAATACTGTGATGATGATGATGATGATGATGATGATG 1159
Db 513 TCGGCTGAATTTTATTTCGAAAAGGTCATTTGATTCATCTGATACAC 561

```

RESULT 5
B2548722 1063 bp DNA linear GSS 17-DEC-2002
LOCUS pasci-60_1377.s1 pasci-60 Pseudomonas aeruginosa genomic clone
DEFINITION pasci-60_1377, genomic survey sequence.

ACCESSION B2548722
VERSION B2548722.1 GI:27152303
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1063)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
JOURNAL
COMMENT Contact: Chris K. Raymond

Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source location/Qualifiers

1..1063
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pasci-60_1377"
/clone_id="pasci-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."

BASE COUNT 202 a 316 c 335 g 209 t 1 others
ORIGIN

Query Match 3.8%; Score 81.2; DB 29; Length 1063;
Best Local Similarity 49.8%; Pred. No. 5.2e-12;
Matches 206; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```

Qy 1449 CACGCTGGGCGCGCAGAGAGCTGGCTTTGGCGCGTGGGATTTGCTGGGCGGAGCT 1508
Db 58 CATCCCGGCTCCGAGAGGCTATCGCAAGGCTCTGCTGATGAAATGCGCAAGCTT 117
Qy 1509 AAACCTGCGATCGTGTGCATCATCGACACCTCCGCGCGCAATTCGAGGCGGCTGA 1568
Db 118 CAAGATCCGATCCTTCATTCATCGACAGCCGCGGCTTACCCGGGATGATGCGGA 177
Qy 1569 GGAAGTGGGATCGCAAGCTCGATGCGGCGCACTTGTCCAACTTATGACGCTCCCT 1628
Db 178 GGAAGCGGCGCAAGCGAGCGGCTGGAACCTGCGGATGAGCGGACTGAAGAC 237
Qy 1629 CCCACCGTTTGGTCAATTAATGTCAGGCGCTTGGCGCGCTGCGCCATGCTGCC 1688
Db 238 GCCGATATGCGCACCGTATCGCGAGGCGGCTCCGCGCGCTGCGCATCGGTGT 297
Qy 1689 CGCGATCTGATCTTACGCGGCGCAAAAGCGTGGCTGTCCGATTCGACCAAGGCGC 1748
Db 298 CTGCGACCAATGGAACATGCTGCAATCTCACCTATTCGATGATTCGCGGAAGATG 357
Qy 1749 CTGCGCATCTCTTCCGCAACCAACGCGCGGGAATGATGAGGACAGAGCGT 1808
Db 358 CGCTCCATCTCTGGAAGACCGCGGAGAGCGCGGAAGCCGCGGCGCATGCGCAT 417
Qy 1809 GCAAGCGCACGCACTTTTAAGCCAAAGGCTTATCGACGGGATGTCGCCGAAC 1862
Db 418 CACGCGCGAGCGGCTGAAAGGCTGCGGCATGTCGACAAAGTCAATGACGAAC 471

```

RESULT 6
B1310727 712 bp mRNA linear EST 20-JUN-2001
LOCUS B1310727
DEFINITION EST5132477 GSSD Medicago truncatula cDNA clone pGSPB024 5' end,
mRNA sequence.

ACCESSION B1310727
VERSION B1310727.1 GI:14985054
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.

REFERENCE 1 (bases 1 to 712)
Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Uteerback,T., Cho
J. and Frazer,C.M.
ESTs from developing reproductive tissues of Medicago truncatula
Unpublished
JOURNAL
COMMENT Contact: Michael A. Grusak

USDA/ARS Children's Nutrition Research Center
Baylor College of Medicine
1100 Bates Street, Houston, TX 77030-2600, USA
Tel: 713-798-7044
Fax: 713-798-7078
Email: mgyrusak@bcm.tmc.edu
3397582e

TIGR sequence name: MTPAL60TX
More information is available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gtc gat CC).

FEATURES

```

1..712
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3890"
/clone="pGBSBJ24"
/tissue_type="immature seeds"
/dev_stage="Immature seeds, 11 to 19 days after
pollination"
/clone_id="GBSD"
/notes="vector: pBluescript SK-, Site_1, EcoRI, Site_2:
XhoI; Immature seeds, collected from pods ranging in age
from 11 to 19 days after pollination, were harvested from
greenhouse-grown plants. Seeds were removed and
separated from pod walls and were immediately frozen in
liquid nitrogen. Seeds throughout the age range were
pooled for mRNA extraction. cDNA was prepared from polyA+
mRNA. The cDNA was directionally ligated into
the UniZap XR vector from Stratagene and packaged using
GigaPack III Gold packaging extracts. Plasmids containing
cDNA inserts were excised from the recombinant lambda-zap
phage using Ex-assist helper phage and propagated in
E. coli cells."

```

BASE COUNT	222 a	107 c	153 g	230 e
ORIGIN				

Query Match	3.8%;	Score 81;	DB 12;	Length 712;
Best Local Similarity	48.4%;	Pred. No. 4.8e-12;		
Matches 256;	Conservative	0;	Mismatches 270;	Indels 3;
				Gaps 1

QY	634	AATCAAGGCTTNGCAGAGACCTTGGAGCGGGCTCGAAGCAAGGCCAAATGCGATGAATGCG	693
Db	63	AATGATGATTTACAAATTCGTCTTATCTTATCAAGACAGAACTGGAATTACTGGAATGCG	122
QY	694	GTAATTACTGAGAAAGGCAACCGTGGAGGGCAATCCGGTAACCGGTATTTTGTCCGATTTT	753
Db	123	GTTCAAAACGACACAGGTCGAAGTAATGGTATTTCCGTAGCAATAGTATTATGATTTT	182
QY	754	TCCTTCCCGGCGGTTCTTTGGGCAACGGTCGGTGGCGGATATATGAAGCCATTCAAC	813
Db	183	GAGTTTATGGAGGTAATGATGGATCCGTATGGGTGAGAAATCACTCGGTTGATTGAA	242
QY	814	CGCGCACAAGAGCTGAAATCCCAATGCTGGCTCCCTCGCTTCGGTGGTGGCGGANG	873
Db	243	TATGTCNCCATCAACGTTTACCTTTATATAGTATGTGCGTCTGGAGAGACGGCTATG	302
QY	874	CAGGAGACAAATCGAGCTTTTGTCTATGATGTGTCCATPACCGCGGCTGTGCAAGGTCAC	933
Db	303	CMAAGAGAAATTGAGCTTATATGCAAAATGGCTAAATTTCTGCTTCTTATATATTAAT	362
QY	934	CGCAGGCGCATTTGCCGTTCTCGTGTATTTGGGCA---TCCAGCATGGGTGGCGCC	990
Db	363	CAATTCATCAAAAGTTATTTCTATGTACCGATACTTAATCTCTACTACTGTGTGGGTA	422
QY	991	ATGACCTCTGGGGTTCATCTGGGCATCTCACTTTTGGGAACCGGCGCGCGAATAGT	1056
Db	423	ACACTAGTTTGGATGTTGGGGGATATCAATTTTCCGAACCCGATGCTTAACATACCA	482
QY	1051	TTCTCGGTCTCTCGGTGTGGAGTTAACCACTGGGCATGCGCTTCCAGAGGTGTGAG	1110
Db	483	TTTCAGGTAAGATGATTAAGAGAAACGTTGAAGATCGAAGTCCCGAAGGTATACAA	542

Qy 1111 CAGCGGAGAAATTGGTGAACACTGCTGATTCATGGAATTGTCGC 1155

Db 543 TCGCGTAATTTTATTGAAAAGGCTGATTTGATTCACTCGTACAC 591

RESULT 7	CA990607	580 bp	mRNA	linear	EST 06-JAN-2007
LOCUS	CA990607				
DEFINITION	CA990607	580 bp	mRNA	linear	EST 06-JAN-2007
	EST644115	GenSeq	Medicago truncatula	cdna	clone GESP-29C10, mRNA
ACCESSION	CA990607				
VERSION	CA990607				
KEYWORDS	CA990607.1	GI:27523501			
SOURCE	EST.				
ORGANISM	Medicago truncatula (barrel medic)				
	Medicago truncatula				

REFERENCE 1 (bases 1 to 580)
AUTHORS Grusak, M.A., Ssmac, D.A., Town, C.D., Van Aken, S., Uterback, T.,
TITLE Cheung, F. and Fraser, C.M.
More ESTs from developing reproductive tissues of Medicago
truncatula
JOURNAL Unpublished
COMMENT Contact: Grusak, M.A.

TIGR sequence name: MTPBX17TK
More information is available at: www.medicago.org
Seq primer: Skmod (CTA GAA CTA gtc gat CC).

```

/organism="Medicago truncatula"
/mol type="mRNA"
/cultivar="A17"
/doi_xref="taxon:3880"
/clone="GESD-29C10"
/tissue type="immature seeds"
/dev stage="immature seeds, 11 to 19 days after
pollination"
/clone_idb="GESD"
/note="Vector: pbluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Immature seeds, collected from pods ranging in age
from 11 to 19 days after pollination, were harvested from
greenhouse-grown plants. Seeds were removed and
separated from pod walls and were immediately frozen in
liquid nitrogen. Seeds throughout the age range were
pooled for mRNA extraction. cDNA was prepared from polyA+
enriched RNA. The cDNA was directionally ligated into
the Unizap XR vector from Stratagene and packaged using
GigaPack III Gold packaging extracts. Plasmids containing
cDNA inserts were excised from the recombinant lambda-Zap
phage using Ex-assist helper phage and propagated in
XIOLR cells."

```

BASE COUNT	171 a	88 c	129 g	192 f
ORIGIN				

Query Match	3.8%;	Score 79.8;	DB 14;	Length 580;
Best Local Similarity	49.5%;	Pred. No. 9.5e-12;		
Matches 235; Conservative	0;	Mismatches 237;	Indels 3;	Gaps 1

QY 688 GAAATCGGTAATTAATCTCGAGAGAGGACCGTSGAGGGCAATCCGGATAGCCGTAATTTGTC 747
Db 2 GATGGCGTTCAACACGGACACAGCTCACTAAATGTAATCTCTGTGACAAATAGGTAATTATG 61
QY 748 GATTTTTCCTCTCGCGCGTTCTTTGGCACGCGTGCCTGCTGCGCATCATGAAAGCG 807
62 GATTTTGAATTATAGGAGAGTAGTATGGGATCCCTAGTGGGTAGAAATCACTCGGTTG 121

BASE COUNT	264 a	112 c	151 g	265 t
ORIGIN				
Query Match	3.8%	Score 79.8	DB 10	Length 792
Best Local Similarity	49.5%	Pred. No. 1.1e-11		
Matches 235	Conservative 0	Mismatches 237	Indels 3	Gaps 1
OY	688	GAATCCGTAATTACTGTGAGAAGGACACCGGTGAGGGCATTCGCGTAGCCGTTATTTTGCC	747	
DB	33	GATCGGGTTCAAAACAGGACACGGTCAAGTAATGTATTCCTGTAGCAATAGGATTAATG	92	
OY	748	GATTTTTCCTTCCTCCGGCGGTTCTTTTAGGACACGGTCCGCGTGGCCATCATGAAGCG	807	
DB	93	GATTTTAGTTTAGTTAGGAGGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG	152	
OY	808	ATTCAACCGCGCACAGAGCTGAAATCTCCACTGCTGTCTTCCTCGTTCCGGTGTGCG	867	
DB	153	ATTGAATATGTGTACCAATCAACCTTTACCTTTTATATGATGTGGTCTGGAGGAGCG	212	
OY	868	CGCATGACAGGAAGCAATTCAGACTTTTGTATGATGATGATGATGATGATGATGATGATG	927	
DB	213	CGTATGCAAGAGGAGAGTTGAGCTTATGCAATATGCTAAATTTCTGCTTCTTTATAT	272	
OY	928	CGTACCCGCGAGGGCGCATTTGCGCTTCCTGCTGTATTTGCGCA--ATCCACAGTGGT	984	
DB	273	AATTATCAATATCAATCAATCAAAAGTTATCTATGATGACGATACATCTCTACTAGTGT	332	
OY	985	GGCGCCATGCGCTGTGGGGTTTCACTCGGCATCTCACTTTTGGGAAACCGCGCGCAG	1044	
DB	333	GGGGTAAACAGCTAATTTTGAATGTGGGGGATATCAATCTTGGCGAACCCGATGCTTAC	392	
OY	1045	ATAGATTTCCTGGGTCCTCCGCGTGTGAGTAAACCACTGGGCATGCGCTTCCAGACGT	1104	
DB	393	ATTACATTGGAGGTAAAGATTAATGGAAGAAACGTTGAAGATCGAAGTCCGGAAGGT	452	
OY	1105	GTGACAGACGGCGGAATTGTGTCAAAACTGTGTGTATTTGATGGAATTTGTCTGC	1159	
DB	453	ATACAAATCGCGTGAATTTTATTCGAAAAGGCTGCAATTGATCATCTGTACAC	507	
RESULT 9				
LOCUS	BZ561247	1042 bp	DNA	linear
DEFINITION	pac62-164_3167.y3 pac62-164 Pseudomonas aeruginosa genomic clone			
ACCESSION	pac62-164_3167			
VERSION	BZ561247			
KEYWORDS	BZ561247.1	GI:27180878		
SOURCE	GSS.			
ORGANISM	Pseudomonas aeruginosa			
REFERENCE	Pseudomonas aeruginosa			
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.			
TITLE	1 (bases 1 to 1042)			
JOURNAL	Spencer D.H., Raymond C.K., Smith E.E., Sims E.E., Hastings M., Burns J.E., Kaul R. and Olsen M.V.			
COMMENT	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press			
	Contact: Chris K. Raymond			
	Genome Center			
	University of Washington			
	Box 352145, Seattle, WA 98105-2145, USA			
	Tel: 2062216954			
	Fax: 2066857244			
	Email: craymond@u.washington.edu			
	Class: Shotgun.			
FEATURES	location/Qualifiers			
SOURCE	1..1042			
	/organism="Pseudomonas aeruginosa"			
	/mol_type="genomic DNA"			

/strain="2-164"
/db_xref="taxon:287"
/clone="pac82-164.3167"
/note="clinical isolate 2-164 whole genomic shotgun library."
BASE COUNT 207 a 320 c 270 g 244 t 1 others

Query Match 3.7%; Score 78.8; DB 29; Length 1042;
Best Local Similarity 50.3%; Pred. No. 2.6e-11;
Matches 194; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 1477 GCGCTGCGGCTATTCGCTGCGGCGGCGAGCTAAACCTGCCATCTGTCATCATCGAC 1536
DB 110 GCTGCGGCTATTCGCTGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 169
QY 1537 ACCTCGGCGGCGGCTATTCGCGAGCGGCTGAGAGCTCGGCAATCGCATTCGCG 1596
DB 170 ACGCCCGGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 229
QY 1597 GCGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1656
DB 230 TGGAACTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 289
QY 1657 GCGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1716
DB 290 GCGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 349
QY 1717 GCGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1776
DB 350 TCCACCTATTCGCTGCTGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 409
QY 1777 CACGCGGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1836
DB 410 AAGGCGGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 469
QY 1837 CTTATCGAGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1862
DB 470 ATCTGCGAGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 495

RESULT 10
LOCUS B1725661 539 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031080C09.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION B1725661
VERSION B1725661.1 GI:15701356
KEYWORDS EST
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 539)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,
P., McDermott, J. P., Shlager, D., Sillitow, C. and Stern, D.
TITLE Analysis of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
JOURNAL Unpublished
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu

FEATURES
SOURCE 1. 539
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21g"

/db_xref="taxon:3055"
/clone.lib="C. reinhardtii CC-1690, Stress II (normalized)
), Lambda Zap II"
/note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into Lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with Exsist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldi et al., (1996) Genome
Research 6: 791-806."

BASE COUNT 95 a 185 c 175 g 84 t

Query Match 3.7%; Score 78.4; DB 12; Length 539;
Best Local Similarity 51.0%; Pred. No. 2.3e-11;
Matches 210; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

QY 1449 CACGCTTGGCGCGGCGGAGCTGCTTGGCGGCTGCGCATTCGCTGCGGCGGAGCT 1508
DB 56 CATGCCCGGCGGCGGCGGAGCTGCTTGGCGGCTGCGCATTCGCTGCGGCGGAGCT 115
QY 1509 AAACCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1568
DB 116 TGGTCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 175
QY 1569 GAGGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1628
DB 176 GAGGCTGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 235
QY 1629 CCCACCTGCTGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1688
DB 236 GCCCATCATTCGCTGCTGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 295
QY 1689 GCGGCTATTCGCTGCTGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1748
DB 296 GCGGCTATTCGCTGCTGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 355
QY 1749 CTCGCGCATTCCTTTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1807
DB 356 GCGGCGCATTCCTTTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 415
QY 1808 TCGAGGCGGCGGCTATTCGCGGCGGCGAGCTTCAAGATCGCTTCACTTCATCGAC 1859
DB 416 TCACCTGCGGCGGAGCTGCTGAGGCTTCAAGATCGCTTCAAGATCGCTTCACTTCATCGAC 467

RESULT 11
LOCUS AG247247 652 bp DNA linear GSS 13-DEC-2002
DEFINITION Ag247247 Locus japonicus DNA, clone: LJT20K03_not, genomic survey sequence.
ACCESSION AG247247
VERSION AG247247.1 GI:26647012
KEYWORDS GSS.
SOURCE Locus japonicus
ORGANISM Locus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotetea;
Locus.
REFERENCE 1
AUTHORS Sato, S., Nakamura, Y. and Tabata, S.
TITLE Locus japonicus TAC End sequences
JOURNAL Published only in Database (2002)
REFERENCE 2 (bases 1 to 652)

AUTHORS		TITLE		JOURNAL		FEATURES	
Sato, S.		Direct Submission		Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail: ssato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/), Tel: 81-438-52-3935 (ex. 2336), Fax: 81-438-52-3934)		Location/Qualifiers	
SOURCE		1. 652					
		/organism="Lotus japonicus"					
		/mol_type="genomic DNA"					
		/strain="Miyakojima Mg-20"					
		/db_xref="taxon:34305"					
		/clone="Uj20k03 not"					
		/clone_lib="genomic PAC library"					
		/note="VECTOR: pYLtAC7"					
BASE COUNT		198 a 107 c 129 g 218 t					
ORIGIN							
Query Match		3.7%; Score 77.6; DB 29; Length 652;					
Best Local Similarity		50.0%; Pred. No. 4, 5e-11;					
Matches 222; Conservative		0; Mismatches 219; Indels 3; Gaps 1;					
Qy	719	AGGGCATTCGGGTACCCGTTATTTTGTCCGATTTTCTCCTCGCGCGTTCTTGGGCA	778				
Db	72	ATGTAATTCCTGTACCAATTGGGATTTAGATTTGAAGTTTATGGAAGTATGATGGAT	131				
Qy	779	CGGTGCGGTGGTGGCATCATGAAAGGCGATTCCAGCGCCACAGAGCTGAAACTCCGAC	838				
Db	132	CCGTGATGATGAGTGAAGAAATCACCCGGTTGGTGAATATGCTACCAACCACTTTTACCTC	191				
Qy	839	TGCTGCTCTCCCTCTCTCTCCGTTGGTGGCGGATCAGAGAAACAATGAGCTTTTGCA	898				
Db	192	TTATTGTATGATATGTCATCTGGAGAGACAGTATCAAGAGAAAGTTTGAATTGAATGC	251				
Qy	899	TGATGATGTCATTAACCGCGGCTGTGCAGCGCTCACCGGAGGCGCATTTGCCGTTCTGTG	958				
Db	252	AAATGGCTAAATATATCTTCTGCTTTATATGATTTATCACTAAATATAAGTTATTCTATG	311				
Qy	959	TGTATTTTGGCGCAA--TCCCAAGATGGGTGGGCCCATGAGCTTCGTGGGTTTCACTGGGC	1015				
Db	312	TATCAATCTTATACATCTCCCACTACTGATGGGGTAAACGATAGTTTGACATGTTGGGGG	371				
Qy	1016	ATCTCACTTTTGGCGGAACCCGGCGCGCAATATGATGTTTCTGGGTCTTCGCGTGTGAGT	1075				
Db	372	ATATCATATTATGGCCGAACCAATATGCTTCAATTCATTTTCGGGGTAAAGAGTAATTGAAC	431				
Qy	1076	TAAACACATGGGATCGGCTTCAGACGGGTGTGCACAGCGGAGAAATTGGTGAATAACTG	1135				
Db	432	AAACCTTGATTAAGCAGATCTCCGGAAGTTACACAGCGCGAGAAATTTTATTCATTAAG	491				
Qy	1136	GTGTGATTGATGATGATTTGTGTGC	1159				
Db	492	GCTGTGTTGATTCATCGTACAC	515				
RESULT 12							
CB977119/c		617 bp mRNA linear EST 01-MAY-2003					
LOCUS		CB977119					
DEFINITION		CA440003_IVA_Ra_D11 Cabernet Sauvignon Berry - CA44 Vitis vinifera					
ACCESSION		CB977119					
VERSION		CB977119.1 GI:30300405					
KEYWORDS		EST.					
SOURCE		Vitis vinifera					
ORGANISM		Vitis vinifera					
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids					
AUTHORS		1 (bases 1 to 617)					
		Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook					
D.							
TITLE		Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'					

JOURNAL
unpublished
Contact: Douglas Cook, PhD
CAES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcoc@ucdavis.edu
Seq primer: GCCAACGAATGCTCAG.
Location/Qualifiers

FEATURES
source
1..617
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultiivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB40003_Iva_Ra_D11"
/sex="Hermaphrodite"
/dev_stage="Berry on stage II, 9 mm"
/lab_host="DHalpha"
/clone_lib="Cabernet Sauvignon Berry - CAB4"
/note="Organ: Berry; Vector: pDNR; Site_1: Sfil; Site_2:
SfiI; CAB4 is a cDNA library of Vitis vinifera cv.
'Cabernet Sauvignon'. Clone 8 berries. Sampled berries were
collected from field-grown vines during stage II of berry
growth (berries were green and hard) at approximately 60
days after full bloom. The average berry size was 9
millimeters. Sampled vines were located at the University
of California, Davis, Experimental Vineyard. cDNAs were
made by oligo-dT priming and directionally cloned. 5' and
3' adaptors were used in cloning as follows:
5'-AAGCATGTGTTTCACCGCAGATGCGCCATTACGGCCGG-3' and
5'-ATTCTAGAAGGCCGAGCGCGCCGACATG-dt(30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size-selected to contain the 0.5-3 kb size fraction."

BASE COUNT
191 a 129 c 94 g 203 t

ORIGIN

Query Match 3.5%; Score 75; DB 14; Length 617;
Best Local Similarity 47.8%; Pred. No. 2.5e-10;
Matches 250; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

Dy 640 GGCTATGCAGAGACTCTTGAGCGGGCTTCGAAGCAGGCCAAATCCGATGATCGGTAA 699
|||
Db 616 GGTATAAAGGCGGTATGATTTTATCAAAGAAAGCAGGATTAAGTAGGCGCTTCAA 557
|||
Qy 700 ACTGAGAGAAGCACCGTGAAGGGGATTCGGTAGCCGTTATTTGTCCGATTTTTCTTC 759
|||
Db 556 ACAGGACACAGGTCACTTAACGAGTATCCGCTGACAATTGGGGTTATGATTTTCAGTT 497
|||
Qy 760 CTCGGCGGTTCTTTGGGACCGGTGCGCTCGGTGGCATCAAGAAGCGATTACCGCGCC 819
|||
Db 496 ATGGGGGGTAGTAGGAGTCGATCGTAGTAGAGAAAATTACCCTTGATGGAATATGCC 437
|||
Qy 820 ACAAGACTGAATCTCCCACTGCTGCTCCCCCTTCCTCCGTTGCTGGTGTGGCGCATCGAGAA 879
|||
Db 436 ACCAATGAATTTCAACCTCTTATTTTGTGTGTCTTCCGAGAGGACGATNCAAGAA 377
|||
Qy 880 GACATGAGCTTTTTCATGATGATGATGATCCATTAACCGCGGTGCGACGCGCAACGGAG 939
|||
Db 376 GGAAGTTTGAAGCTTGATGACAAATGGCTTAANAATATCTTCCTGTTATATATGATTAATCA 317
|||
Qy 940 GCGCATTTGCGCTTCCTGATGATTTTGGCAA--TCCACGATGGGTGGCGCCATGACC 996
|||
Db 316 AATAAAGATTTATTTATGATATCAATCTTAATCAATCTCTACTACTGATGGGGGTACAGCA 257
|||
Qy 997 TCGTGGGGTTCATGTGGGCATCTCATCTTTTGGGAACCGCGGCGCGCAATAGTTTCTTG 1056
|||
Db 256 AGTTTGGTATGTGGGAGATATCATATATATGGCGAACCCTAATTCCTCATTTGCACTTGGC 197
|||
Qy 1057 GGTCTCGCGGTGGAGAGTTAACCACTGGGGCATCGCTTCAACGCGGTGTGACAGAGCG 1116
|||
Db 196 GGTAAAAAGGTATATGAACAAACATTGAAAAAGCACTACTCTGAAGGTACCAAGCGGCT 137
|||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:57:27 ; Search time 131.654 Seconds
(without alignments)
7117.550 Million cell updates/sec

Title: US-10-024-370-1

Perfect score: 2123

Sequence: 1 ctccgagcgsgagctcgatgcgtatc.....gacctcatgcctggaattc 2123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	2123	4 US-09-362-899-1	Sequence 1, Appl1
2	1473	66.4	1473	4 US-09-362-899-2	Sequence 2, Appl1
3	312.6	14.7	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
4	312.6	14.7	4411529	3 US-09-103-840A-1	Sequence 1, Appl1
5	142	6.7	36063	4 US-08-311-731A-140	Sequence 140, App
6	116.4	5.5	1020	4 US-09-252-991A-12180	Sequence 12180, A
7	116.4	5.5	1338	4 US-09-252-991A-12415	Sequence 12415, A
8	108.6	5.1	23673	4 US-09-773-816-1	Sequence 1, Appl1
9	104.2	4.9	870	4 US-09-107-532A-170	Sequence 170, App
10	97.2	4.6	444	4 US-09-252-991A-12365	Sequence 12365, A
11	96.2	4.5	19702	4 US-08-961-527-7	Sequence 7, Appl1
12	96.2	4.5	1830121	4 US-09-557-884-1	Sequence 1, Appl1
13	96.2	4.5	1830121	4 US-09-643-990A-1	Sequence 1, Appl1
14	93.2	4.4	2561	4 US-08-347-878-25	Sequence 25, Appl1
15	87.4	4.1	954	4 US-08-973-275-2	Sequence 2, Appl1
16	86.4	4.1	903	4 US-09-328-352-1114	Sequence 1114, Ap
17	85.6	4.0	1230025	4 US-09-198-452A-1	Sequence 1, Appl1
18	84.8	4.0	954	2 US-08-973-275-3	Sequence 3, Appl1
19	84	4.0	10303	4 US-09-634-238-410	Sequence 410, App
20	79.6	3.7	999	4 US-09-252-991A-4651	Sequence 4651, App
21	77.8	3.7	1203	4 US-09-252-991A-4229	Sequence 4229, Ap
22	77.2	3.7	456	4 US-09-134-001C-366	Sequence 12319, A
23	75.2	3.5	882	4 US-08-714-918-105	Sequence 966, App
24	72.8	3.4	2255	3 US-09-265-315-105	Sequence 105, App
25	72.8	3.4	2255	3 US-09-265-315-105	Sequence 105, App
26	72.8	3.4	2255	3 US-09-266-417-105	Sequence 105, App
27	72.8	3.4	2255	3 US-09-266-417-105	Sequence 105, App

28	54.6	2.6	3183	1 US-08-849-212-3	Sequence 3, Appl1
29	51	2.4	2894	1 US-08-278-091-1	Sequence 1, Appl1
30	51	2.4	2894	1 US-08-483-851-1	Sequence 1, Appl1
31	51	2.4	2894	1 US-08-472-173-1	Sequence 1, Appl1
32	51	2.4	2894	2 US-08-487-167-1	Sequence 1, Appl1
33	51	2.4	2894	2 US-08-482-816-1	Sequence 1, Appl1
34	51	2.4	2894	2 US-08-296-149-1	Sequence 1, Appl1
35	51	2.4	2894	2 US-08-801-499-1	Sequence 1, Appl1
36	51	2.4	2894	2 US-08-615-271-1	Sequence 1, Appl1
37	51	2.4	2894	3 US-09-074-660-1	Sequence 1, Appl1
38	51	2.4	2894	3 US-09-074-650-1	Sequence 1, Appl1
39	51	2.4	2894	3 US-09-106-468-1	Sequence 1, Appl1
40	51	2.4	2894	3 US-09-106-468A-1	Sequence 1, Appl1
41	51	2.4	2894	3 US-09-106-467-1	Sequence 1, Appl1
42	49.8	2.3	789	4 US-09-107-532A-1183	Sequence 1183, Ap
43	48.2	2.3	984	1 US-08-611-107-11	Sequence 11, Appl
44	48.2	2.3	984	2 US-08-422-560A-11	Sequence 11, Appl
45	48.2	2.3	984	3 US-08-468-793-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1									
US-09-362-899-1									
; Sequence 1, Application US/09362899									
; Patent No. 6361986									
; GENERAL INFORMATION:									
; APPLICANT: Degussa-Hls AG									
; APPLICANT: Forschungszentrum-Jlich GmbH									
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND									
; CURRENT APPLICATION NUMBER: US/09362,899									
; EARLIER FILING DATE: 1999-07-29									
; EARLIER APPLICATION NUMBER: DE 19924365.4									
; EARLIER FILING DATE: 1999-05-27									
; NUMBER OF SEQ ID NOS: 3									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 1									
; LENGTH: 2123									
; TYPE: DNA									
; ORGANISM: Corynebacterium glutamicum									
; FEATURE:									
; NAME/KEY: gene									
; LOCATION: (508)..(1980)									
; OTHER INFORMATION: accda									
US-09-362-899-1									
Query Match									
Beet Local Similarity 100.0%; Pred. No. 0; Length 2123;									
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	CTCGAGCGGAGTGGTATCGGCACCTCTTAAGCAATGCGCGCTTAAATAAGCA	60						
DB	1	CTCGAGCGGAGTGGTATCGGCACCTCTTAAGCAATGCGCGCTTAAATAAGCA	60						
QY	61	CTTAATGTTTCTACCACTCTGGCGACGACGACGAGATGTTGATCAGCTA	120						
DB	61	CTTAATGTTTCTACCACTCTGGCGACGACGACGAGATGTTGATCAGCTA	120						
QY	121	AACGGTAAGTGAATGATCTTAACCTAAGCAATGCAATGCGATTTGGAATCTC	180						
DB	121	AACGGTAAGTGAATGATCTTAACCTAAGCAATGCAATGCGATTTGGAATCTC	180						
QY	181	ACTGCCCAATATCTTAACCTAACTTAAGTATGTTTACCTGATTTAAAGT	240						
DB	181	ACTGCCCAATATCTTAACCTAACTTAAGTATGTTTACCTGATTTAAAGT	240						
QY	241	TCCGATTTACCCCTCTTTACCCGGAATATCCCTTTTGCAGAAATGCAACACA	300						
DB	241	TCCGATTTACCCCTCTTTACCCGGAATATCCCTTTTGCAGAAATGCAACACA	300						
QY	301	GTGCAATGATTAAGGGGCTTACAGCAGTCAATCTGCGGTTTAGCTATGTTGGG	360						
DB	301	GTGCAATGATTAAGGGGCTTACAGCAGTCAATCTGCGGTTTAGCTATGTTGGG	360						

301 GTGCAATAGTTAACGGGCTTCACAGCTCACCATTCGTCCGGTTTATAGTTAGTTCCGG 360
361 ACGTCTAGGCAAAAGTAGTTTGTGAGATGAAAGCAATCCGTCATTTTATTCGGA 420
421 ACGTCTAGGCAAAAGTAGTTTGTGAGATGAAAGCAATCCGTCATTTTATTCGGA 420
421 TCGATAGCTAAATTTGGCTTATGATCTTCGCTCTTAATATGATGACAGACATTCGA 480
481 TCGATAGCTAAATTTGGCTTATGATCTTCGCTCTTAATATGATGACAGACATTCGA 480
481 ATTAATTAACAAGCAATTTTCCGCTCGTGAAGAACGCTTTCCGATATGATGATGAGG 540
481 ATTAATTAACAAGCAATTTTCCGCTCGTGAAGAACGCTTTCCGATATGATGATGAGG 540
541 ATGAACACACTTCGATGATGACCTGATGACCTGATGATGATGATGATGATGATGATG 600
541 ATGAACACACTTCGATGATGACCTGATGACCTGATGATGATGATGATGATGATGATG 600
601 TCTTGAATGAAACTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660
601 TCTTGAATGAAACTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660
661 CGGCTCGAAGCAAGGCAATGCGATGATGATGATGATGATGATGATGATGATGATG 720
661 CGGCTCGAAGCAAGGCAATGCGATGATGATGATGATGATGATGATGATGATGATG 720
721 GGCATTCGAGTACCGGCTATTTGTCGATTTTCTTCGATTTTCTTCGATTTTCTTCG 780
721 GGCATTCGAGTACCGGCTATTTGTCGATTTTCTTCGATTTTCTTCGATTTTCTTCG 780
781 GTCGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
781 GTCGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
841 CTGCTCTCCCTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
841 CTGCTCTCCCTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
901 ATGGTGCATTAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
901 ATGGTGCATTAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
961 ATGGTGCATTAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
961 ATGGTGCATTAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
1021 ACTTTTGGGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1021 ACTTTTGGGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1081 ACTGGGCAATGCGCTTCCAGACGATGATGATGATGATGATGATGATGATGATGATG 1140
1081 ACTGGGCAATGCGCTTCCAGACGATGATGATGATGATGATGATGATGATGATGATG 1140
1141 ATTGAATGATTTGCTGCTCACTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
1141 ATTGAATGATTTGCTGCTCACTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
1201 ATTGAATGATTTGCTGCTCACTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
1201 ATTGAATGATTTGCTGCTCACTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
1261 GTGATGAGAGCGATTTGCGGCTTCTGATGACCGGCAAGGCTGATGATGATGATGATG 1320
1261 GTGATGAGAGCGATTTGCGGCTTCTGATGACCGGCAAGGCTGATGATGATGATGATG 1320
1321 GAAACGTTGGGAGGAGAGAGTGTCTGATGATGATGATGATGATGATGATGATGATG 1380
1321 GAAACGTTGGGAGGAGAGAGTGTCTGATGATGATGATGATGATGATGATGATGATG 1380
1381 GCTGTGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
1381 GCTGTGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

1441 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1441 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1501 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
1501 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
1561 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
1561 GCGGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
1621 GCTCCCTCCCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1621 GCTCCCTCCCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1681 ATGCTGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1681 ATGCTGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1741 GAGGCGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1741 GAGGCGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1801 CAAGGCTGAGAGGCGAGCTTTTAAAGCAAGGCTTAAAGCAAGGCTTAAAGCAAGGCT 1860
1801 CAAGGCTGAGAGGCGAGCTTTTAAAGCAAGGCTTAAAGCAAGGCTTAAAGCAAGGCT 1860
1861 ACCGAGCACTTTGTTGAAGAAATTTCTGCGACATGACAAACGCTTCTCGAATGGAT 1920
1861 ACCGAGCACTTTGTTGAAGAAATTTCTGCGACATGACAAACGCTTCTCGAATGGAT 1920
1921 AACCAATCGGAGAGGCGGAGAGCGGACATGCTTCAACAGATTTGAAGCTTTAGCGAG 1980
1921 AACCAATCGGAGAGGCGGAGAGCGGACATGCTTCAACAGATTTGAAGCTTTAGCGAG 1980
1981 TAAAGAAATTTAAGCGCTGATCAATGATGATGATGATGATGATGATGATGATGATG 2040
1981 TAAAGAAATTTAAGCGCTGATCAATGATGATGATGATGATGATGATGATGATGATG 2040
2041 GTGGCGGAAACCTTCAGAGGCGGTAAGAGGCTTGGCGGATGATGATGATGATGATGATG 2100
2041 GTGGCGGAAACCTTCAGAGGCGGTAAGAGGCTTGGCGGATGATGATGATGATGATGATG 2100
2101 GCGGACCTTATGCTTGAATTC 2123
2101 GCGGACCTTATGCTTGAATTC 2123

RESULT 2
US-09-362-899-2
Sequence 2, Application US/09362899
Patent No. 6361986
GENERAL INFORMATION:
APPLICANT: Degussa-HLS AG
APPLICANT: Forschungszentrum-Jlich GmbH
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND
CURRENT FILING DATE: 1999-07-29
EARLIER APPLICATION NUMBER: DE 19924365.4
EARLIER FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentm Ver. 2.1
SEQ ID NO 2
LENGTH: 1473
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1473)
OTHER INFORMATION: accDA
US-09-362-899-2

Qy	1112	AGCGCGAGAAATTGGTGA AAACTGGTGTGATTGATGAAATTGTGTGCGCACTCCAAATTGC	1171
Db	1007607	CCGCCGAGATCTACGGCGGCAGTGGAAATCATGACGGGGTGTGCTGCACTGGAACCGGCTAC	1007548
Qy	1172	GTGCAAGCGGTGGCAAAAACCTCAAGTATAATCACCGGTAGAGGCAACGATCGTTT	1231
Db	1007547	GACCGATGCTGGATGTTGTGTGATTGACGGTCTCATGACGCTCCGAAACCGCTTCCGGCAC	1007488
Qy	1232	CTTCAACAACTCTCGGCGTGGCA---CTTCCGGTATGGAGCGCATTCGCGTTCTCGTG	1288
Db	1007487	CGCAACACCCCGCCGCCGATACCCGATGTGCCCCACCTGGGAGTACGATGTGGCATGCGCC	1007428
Qy	1289	ACCCGCAAGAGCGCTGGAATTCGGGGAGATTATGAAAAGTTGGGGGCAACAGTCTGTAAC	1348
Db	1007427	GCGCCGACCGGGCGGGCGGTCAAGGACGTACTGCGACACGGGCCACCCACCGGGTGTGT	1007368
Qy	1349	TTTCTGTGTCGGCTGTGGCGCATTTAGACCCGGCTGTGCGGTTGCTCCCTGGCGGCATCG	1408
Db	1007367	TGTCAAGAAACGATCAAGGCCAAGG---GCGACCAACGCTGTGCGCTGGCCCGCTTTC	1007311
Qy	1409	GGGGCGGCGCGTGTGCTGATTGGGCGAG-----GATGCGCGCTTCAACG	1453
Db	1007310	GCGGCGAACCCACGGTGTGCTCTCGGCGACGAAAGGACGTAAGGCGGGGGAAGCACTG	1007251
Qy	1454	TTTGGGCGCGAGAGAGCTGTGCTTTTGGCGGCTGTGGGCAATTTCGCTGGCGCGCACTAAAC	1513
Db	1007250	TGGGCGCGCGCTGCTGTAAAGCGAAGCCGACGCGGATGCGCTCGCCGCGAGCTGTGCC	1007191
Qy	1514	TGCCGATGCTGTCAATCATGACACCTCGGCGCGCAATTGTGCAAGCGGCTGTAGAGAC	1573
Db	1007190	TGCGCTGTGTGTGATTTGACGCGCGCGGACCGCGCTGTGCGCGCAACGGAACAGG	1007131
Qy	1574	TGCGCATGCAAGCTGATTTGGCGCGCACTTGTCCAACTTATGACGCTCCCTCCCA	1633
Db	1007130	GCGGCGCTGCGCGGCAGATCGGCGCATTCGCTGGCGCACTGTGATACCGCA	1007071
Qy	1634	CGGTTTGGGTCAATTATGTGTCAAGGCGGTGGCGGCTGTGGCGCTGACATGCTGCCCGCG	1693
Db	1007070	CGGTGTGATCTGTCTGGGCGCAAGGCGACGCGCGCGCTGTGCGATGTTGGCCCGCG	1007011
Qy	1694	ATTGCTTACGCGCGCGCAAAAACGCGTGTGCTGTGCGCATTCGCAACGAGGGGCGCTCGG	1753
Db	1007010	ACCGGCTCTGCGCGCACTTCAAGGCTGTGGCGCGCTTTCCTTCCGAAAGACCAAGCG	1006951
Qy	1754	CCATCTCTTCCGCGACCAACCAACGCGCGGGAATCATATGAGCGCAAGAGCGCTGACAG	1813
Db	1006950	CGATGTGTTCGAGACACTGTCTCATGCGCGCGAATCTGCTGCGCGCCCAAGGATCCGGT	1006891
Qy	1814	CGCAGCAGCTTTTAAGCCAGAGGCTTATGACGCGGATGTGCGCGAAACCGAGCACTTGT	1873
Db	1006890	CGGCGCACTTACTGAAATCGGGGATTTGCAACCATGTGCTCGAGTAACCCGACGCGG	1006831
Qy	1874	TTGAAGA 1880	
Db	1006830	CAGACGA 1006824	
RESULT 5			
US-08-311-731A-140			
Sequence 140: Application US/08311731A			
Patent No. 6583266			
GENERAL INFORMATION:			
APPLICANT: SMITH, DOUGLAS			
APPLICANT: MAO, JEN-I			
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES			
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR			
NUMBER OF SEQUENCES: 411			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.			
STREET: 600 ATLANTIC AVENUE			
CITY: BOSTON			
STATE: MASSACHUSETTS			

```

1  COUNTRY:  USA
2  ZIP:  02210
3
4  COMPUTER READABLE FORM:
5
6  MEDIUM TYPE:  floppy disk
7  COMPUTER:  IBM PC compatible
8  OPERATING SYSTEM:  PC-DOS/MS-DOS
9  SOFTWARE:  PatentIn Release #1.0, Version #1.25
10
11  CURRENT APPLICATION DATA:
12
13  APPLICATION NUMBER:  US/08/311,731A
14
15  FILING DATE:
16
17  CLASSIFICATION:  530
18
19  ATTORNEY/AGENT INFORMATION:
20
21  NAME:  GATES, EDWARD R.
22
23  REGISTRATION NUMBER:  31,616
24  REFERENCE/DOCKET NUMBER:  C0044/7125
25
26  TELECOMMUNICATION INFORMATION:
27
28  TELEPHONE:  617/720-3500
29
30  TELEFAX:  617/720-2441
31
32  INFORMATION FOR SEQ ID NO:  140:
33
34  SEQUENCE CHARACTERISTICS:
35
36  LENGTH:  36063 base pairs
37  TYPE:  nucleic acid
38  STRANDEDNESS:  double
39
40  TOPOLOGY:  circular
41
42  MOLECULE TYPE:  DNA (genomic)
43
44  HYPOTHETICAL:  NO
45
46  ANTI-SENSE:  NO
47
48  ORIGINAL SOURCE:
49
50  ORGANISM:  Mycobacterium leprae
51
52  US-08-311-731A-140

```

Query Match	6.7%	Score 142	DB 41	Length 36063
Best Local Similarity	57.1%	Pred. No. 4.8e-11		
Matches 296	Conservative 0	Mismatches 220	Indels 2	Gaps 2
Qy	685	GATGATCGGTAAATTA	CTGAGAGGACCGGTGAGGGCATTC	CCGTCGGTATTTTG 744
Db	31425	GAGCAATCGGTGCTGAC	CACAGCCAGGGGTGCGTATTCGGGCGAGCGGGTAGCGAATTAATGTC	31488
Qy	745	TCCGATTTTCTTCTCC	TCGGGGGTTCTTTGGGACAGTGCCTGCGTGGCCATCATGAAG	804
Db	31485	TGCAAAATTCAGCTTCT	CTCGGTGGCCGTAATCGGGGTGACACCGGAGCGCAATACCGCC	31544
Qy	805	GCGATTACCGCGCCA	CAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCGGTGTG	864
Db	31545	GCGGTGACGG-GCGA	CAGCTGAGCGGTCGCCG-CGTGGTCTGCGCAAGTTCCGATGGC	31600
Qy	865	GCGCGCATGAGAGAA	CAATCAAGCTTTTGTATATGTGTCTCAATACCGCGCTGTG	924
Db	31603	ACCGGATGAGAGAA	GACACCGGTGCGTCTTCGCAATAGTGAAGATCGTGGCGCGTC	31662
Qy	925	CACGCTACCCGACAG	CGCATTTGCCGCTCTGTGATTTTGGGCATCCACGATGGT	984
Db	31663	AAACTCTCCAAAGG	GCGGGGTCTGCCCTACTTAAGGTACTTGGGTAAATCGACACCGGC	31722
Qy	985	GCGGCATGCGCTCT	GTGGGGTTCATCTGGGCATCTCACTTTTGGGAACCGGCGCGAC	1044
Db	31723	GGGGTCTTCCGTCG	TGGGGCTCTGCTGGGCCAATGAACCGTCGCCACAGCGGGTGTCTCA	31782
Qy	1045	ATAGGTTTCTGTGGT	CTCGCGGTGTGAGATTAAACACTGGGGCATCGCTTCACAGCGGT	1104
Db	31783	ATGGGCTTCTTGGCT	CTCGGGGTATGATGATTTCTGTATAGGGAACCTTTCCGCTCCGAC	31844
Qy	1105	GTCGACGAGCGGAGA	ATTGTGTGAAAACCTGTGATTGATGAATGTGTGCGCACTC	1164
Db	31843	ATCCACACAGCGGAGA	ATCTGCACACGATGGGGTATATGACGCGTATCGTCAAGCTGAC	31902
Qy	1165	CAATTGCGTGAACG	CGGTGCGCAAAAACCCCTCAAGTTAT	1202
Db	31903	GGACTGCAACTGAC	GTTGATCGTGGCTGACGATGAT	31940

US-09-252-991A-12180
; Sequence 12180, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12180
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12180
Query Match 5.5%; Score 116.4; DB 4; Length 1020;
Best Local Similarity 52.0%; Pred. No. 2.7e-24;
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;
QY 686 ATGATCGGTAATTACTGGAAGAAGCACCCTGGAGGCAATCCGGTAAGCCTTATTGT 745
DB 473 AGGAGCGCGTATGCTCCATGAGCGGCAAGCTGCAAGGATGCGGATGCGCTCGCCT 532
QY 746 CCGATTTTCTCTCCGCGCGGTTCTTGAGCAGCGTCCGCTGCGCATCATGAAG 805
DB 533 TCGAGTTCTCTTATGAGCGGTTGATGAGCGCATGTCGGAGAGGCTTCGTTCCG 592
QY 806 CGATTCAACCGCGCCACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGGTGTG 865
DB 533 CAGCCACGTCGCTGAGAGAGCGCTGCGCTGATCTGCTTCGCGCTCCGCGCGG 652
QY 866 CCGCATGAGAGAAACATCGAGCTTTGTATCATAGTGTCCATACCGGCGCTGTC 925
DB 653 CCGCATGAGAGAAAGCGTATGCTGATGCAAGTGGCAAGCCTCGGCGCTCG 712
QY 926 AGCGTCACCGCAGAGCGCATTTGCGTTCCTGATATTGGCAATCCACGATGGTG 985
DB 713 CCGCGCTGCGGAGAAAGGCAATCCGTTCTGCTGATGACCAACCGGCTTAAGGG 772
QY 986 GCGCATGCGCTGCGGCGTATCTGCGCATCTCACTTTGCGAACCAGCGCGCAGA 1045
DB 773 GCGTTCCGCGCAGCTGGGATGCTCGGCGACGTATGCTCGGCAACCCAGGCGCTGA 832
QY 1046 TAGGTTCTCTGCGTCTCGCGTGGAGTTAACACATGGGCGATCGGCTTCAGACGGTG 1105
DB 833 TCGGCTTCCGCGGCTCCCGCGTATGACAGACGCTCCGAGAAAGGCT 892
QY 1106 TGCACAGCGCGAGATTTGTGAAACTGTGTGATGATGAAATTGTGCGCACTCC 1165
DB 893 TCCAGCGTATGAGATTTCTCTTGAAGCATGGGCGCATGACATGATCGATGCGCG 952
QY 1166 AATTGCGTCAAGCGGTGCAAA 1187
DB 953 AGTTGGCGCGCGCGCTGGCCAA 974
RESULT 7
US-09-252-991A-12415/c
; Sequence 12415, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12415
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12415
Query Match 5.5%; Score 116.4; DB 4; Length 1338;
Best Local Similarity 52.0%; Pred. No. 3.1e-24;
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;
QY 686 ATGATCGGTAATTACTGGAAGAAGCACCCTGGAGGCAATCCGGTAAGCCTTATTGT 745
DB 677 AGGAGCGCGTATGCTCCATGAGCGGCAAGCTGCAAGGATGCGGATGCGCTCGCCT 618
QY 746 CCGATTTTCTCTCCGCGCGGTTCTTGAGCAGCGTCCGCTGCGCATCATGAAG 805
DB 617 TCGAGTTCTCTTATGAGCGGTTGATGAGCGCATGTCGGAGAGGCTTCGTTCCG 558
QY 806 CGATTCAACCGCGCCACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGGTGTG 865
DB 557 CAGCCACGTCGCTGAGAGAGCGCTGCGCGATGCTGCTTCGCGCTCCGCGCGG 498
QY 866 CCGCATGAGAGAAACATCGAGCTTTGTATCATAGTGTCCATACCGGCGCTGTC 925
DB 497 CCGCATGAGAGAAAGCGCTGATCTGCTGATGACAGATGGCCAAACCTCGCGCTCG 438
QY 926 AGCGTCACCGCAGAGCGCATTTGCGTTCCTGATATTGGCAATCCACGATGGTG 985
DB 437 CCGCGCTGCGGAGAAAGGCAATCCGTTCTGCTGATGACCAACCGGCTTAAGGG 378
QY 986 GCGCATGCGCTCGTGGGATTCATCTGGGCACTCACTTTGCGAACCAGCGCGCAGA 1045
DB 377 GCGTTCCGCGCAGCGCTGGCATGCTCGGCGACGTATGCTGCGGAAACCCAGGCGCTGA 318
QY 1046 TAGGTTCTCTGCGTCTCGCGTGGAGTTAACACATGGGCGATCGGCTTCAGACGGTG 1105
DB 317 TCGGCTTCCGCGGCTCCCGCGTATGACAGACCGCTCGGAGAAAGGCT 258
QY 1106 TGCACAGCGCGAGATTTGTGAAACTGTGTGATGATGAAATTGTGCGCACTCC 1165
DB 257 TCCAGCGTATGAGATTTCTCTTGAAGCATGGGCGCATGACATGATGCTGATGCTCGG 198
QY 1166 AATTGCGTCAAGCGGTGCAAA 1187
DB 197 AGTTGGCGCGCGCGCTGGCCAA 176
RESULT 8
US-09-773-816-1/c
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; TITLE OF INVENTION: ANTAGONISTS
; FILE REFERENCE: 26600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SEQ ID NO 1
; SOFTWARE: FastSeq for Windows Version 4.0
; LENGTH: 23673
; TYPE: DNA

ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1) (23623)
OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Query Match 5.1%; Score 108.6; DB 4; Length 23673;
Best Local Similarity 46.6%; Pred. No. 2,7e-21;
Matches 611; Conservative 0; Mismatches 619; Indels 81; Gaps 5;

QY 684 CGATGATGCTGATTAATTAAGGAGGACCGTGGAGGGGATTCCGGTAGCCGTTATTTT 743
DB 8617 CGACGAGGCGGCTGTCATGTCACCGCGGATCCACGCGCAGCCGCTGTCGTCGCCG 8558
QY 744 GTCCGATTTTCTCTCTGCGGCTTTTGGCAGGTCGCTCGGTGCGATCATGAA 803
DB 8557 GATGACCTTCGCTTCTGCGGCGGACCTGGCGGCGGCTGGGAACTGATCACTCT 8498
QY 804 GCGCATTCACCGCGCCACAGAGCTGAACTCCCATGCTGTGTCCTCCCTGCTTCCGGTG 863
DB 8497 CCGCCCGGAGACCGCGCTCGCCGACACCTGCGCTGTGATGTGACGCGCTCCGGCG 8438
QY 864 TCCGCGCATGCAAGAAACATGAGCTTTTGTGATGATGTGTCCATAACCGCGCTGT 923
DB 8437 CCGCGCGCATGCAAGAAAGCGCGCTGCGCTCATGACAGATGGCCAGACGAGCGCGCT 8378
QY 924 GAGGCTCACCGCGAGCGCGCATTTGCGCTTCTGCTGTATTTGCGCAATCCACGATGG 983
DB 8377 CCGGCACTGCAAGAGCGCGCGCTGTGACCATCTGCTGCTGACCGACCGACCTTAAC 8318
QY 984 TGGCCCATGAGCTCTGCTGCGGCTTCACTGAGCATCTCACTTTTGGGAAACCGCGCGCA 1043
DB 8317 CCGGCTTCCGGCGCTCTTTCGCACTCTGTGCACTGTCGCGCAACCGCGCGCGCG 8258
QY 1044 GATAGCTTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1103
DB 8257 TCTCGGCTTCCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8198
QY 1104 TGTGACGAGCGGAGAAATTTGGTGAACCTGTGTGATGATGATGATGATGATGAT 1163
DB 8197 GTTCAAGACCGCGCAATACCTCTCTGCAACACGCGCTGATGATGATGATGATGAT 8138
QY 1164 CCAATTGCGTGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1192
DB 8137 CCGGCTTCCGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8078
QY 1193 TCAAGCTTATTCAGCGCTGTAAGGCAACGAT-----CGTTTTTCTCA 1237
DB 8077 GAGCGGGAACCGGCGCGCGCGCTGAGACGCGGCGGCGATGCGCTAGCTGTGCA 8018
QY 1238 CAATCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1297
DB 8017 CCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATGCA 7958
QY 1298 --GGCTTGAATCGGGAAGATTATGAAACGTTGGGGGCAAGCTGTCAACTTTTCG 1355
DB 7957 CCGGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7898
QY 1356 TCCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
DB 7897 GAGCGGCACTGCGGCG 7838
QY 1416 GCGCGTGTGCTGATTTGGCA-----GATGCGCGCTTCAAGCTTGGCGCGAGAG 1467
DB 7837 ACCCGTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7778
QY 1468 CTGCGTTTGGCGCT-----CGTGGCATTTGCGCTGCGCG 1502
DB 7777 CTGCGGCACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7718
QY 1503 CGAGTAAACCTGCGATGCTGTGCTGATGACACTTCCGCGCGCGGATTTGCGAGGC 1562

DB 7717 CAAACTCGGCTCTCCGCTGCTGCACTGTCGACACCCCGGCGCTTACCCGCGCGCT 7658
QY 1563 GCGTGAAGAGCTTCGGATGCAAGCTGCAATTTGCGCGCACTTGTCCAACTTATGACGC 1622
DB 7657 CCGCGAGGAGCAGGCGCAGCGCGCTGCGCATTCGCCGAACTGCGCTTATGCGCGCT 7598
QY 1623 TCCCGTCCCGCACTTTCGCTATTTATGATGAGGCGTGTGCGGCGCGCTGCGCT 1682
DB 7597 GCGCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7538
QY 1683 GCTGCCCGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1742
DB 7537 CCGCGTCCCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7478
QY 1743 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1802
DB 7477 GCGCTGCGGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7418
QY 1803 AGCGTGCAGGCGCGCACTTTTAAGCCAGGCTTATGCAAGGATGCTGCGCGAAC 1862
DB 7417 GAGGCTCACCGCGCGCACTCTGCGCTTGGGCACTGCTGCAAGGCTTCTCCGGAAC 7358
QY 1863 CGAGCACTTTGTTGAAGAAATTCGCGCAATGACCAAGCCCTTCCGA 1913
DB 7357 CGAAGCGCGCACCGGAGCGCGACCGCTGCGCGCGCGCGCGCGCGCTCCGCGA 7307

RESULT 9
US-09-107-532A-170
Sequence 170, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Ducelette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 170:
SEQUENCE CHARACTERISTICS:
LENGTH: 870 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

```

: ORGANISM: Enterococcus faecium
:
: FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (B) LOCATION 1..870
:   SEQUENCE DESCRIPTION: SEQ ID NO: 170
:
US-09-107-532A-170

```

Query Match	4.9%	Score 104.2;	DB 4;	Length 870;
Best Local Similarity	50.9%;	Pred. No. 1e-20;		
Matches 247; Conservative	0;	Mismatches 238;	Indels 0;	Gaps 0;

Qy	686	ATGAATCGGTAAATTCTGAGAAAGGCAACGATGAGAGGCAATTCGCGTACCCGTTAATTTTGT	745
Db	323	ATGAAGCAGTCTCTACGTGTAAAGCAACCATTCAGGGATTCCTTCATATATGGAAGTGA	382
Qy	746	CCGATTTTCTTCTCTCGACGGTCTTTTGGGCACGGTGCGTGGTGGGCATCATGAAGG	805
Db	383	TGGATCTCTAAATTTTATCATGGGAAGTATGGGAACGATTTGCGAAGAAAAGATCACCAAGAT	442
Qy	806	CGATTCACCGGGCCACAGAGCTGAAGTCCCATCTGATCTCCGTCGTTCCGGTGGNG	865
Db	443	TATTGGAACGAGCAACAAAGGAATCTTCCAGTGGTCTTGTTCACAGCTTCAGGTGGCG	502
Qy	866	CGCGCAGSCAGGAAGACATCGAGCTTTTGTCTATGATGGTGTCCATAAACCGCGGCTGTGC	925
Db	503	CACGGATCCAAAGAAAGAAATCTTTTCTTAAATGCATATGGCAAAAATCTCTGTGCTGTCA	562
Qy	926	AGCGTCACCGGAGGCGCATTTTGCCGTTCTCTGTGATTTTGGCGCATCCACGATGGGTG	985
Db	563	AACGGCATAGCAATBAAGTCTTTTATATCTAACGTTTTAAACGAAACCGAGCACCGGAG	622
Qy	986	GCGCCAGTGCCTCGTGGGGTTCATCTGGGCATCTCATTTTGGGAAACCCGGCGCGCAGA	1045
Db	623	GAGTAACTGCGAGTTTTCGAATGGAAGGGAGATATATTCTACCAAAACCCCAAGCTTGA	682
Qy	1046	TAGGTTCTCGGTCCTCGCGCGTGGAGGATTAAACCATGGGGATGCGCTTCCAGACGGTG	1105
Db	683	TGGGTTTGTCTGGAAGACGCGTCATCGAACAGACGATCAAGCAGGAATTTCCAGAAAGATT	742
Qy	1106	TGCAGCAGGCGGAGAATTTGGTGAAGAACTGCTGTATTTGATGAAATTTGTGTGCCACTTC	1165
Db	743	TCCAAAAGGCTGAATTTCTTATATATCTCATAGTGTTTTGTGCATCAGATTTGTTCCCCCAGTGG	802
Qy	1166	AATTG 1170	
Db	803	AATTG 807	

RESULT 10
 US-09-252-991A-12365/C
 ; Sequence 12365, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 10/7196,136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 12365
 ; LENGTH: 444
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-12365

Query Match	4.6%;	Score 97.2;	DB 4;	Length 444;
Best Local Similarity	51.6%;	Pred. No. 8.2e-19;		
Matches 222; Conservative	0;	Mismatches 208;	Indels 0;	Gaps 0

QY	758	TCCTCGGGGCTTTTGGACGGTCCGCTCGGCGGCATCATGAAGGCATTACCGG	817
Db	444	TCATGGCGGTTGCATGGGCGCCATCGTCGGCGAGCGCTTGTTCGGCGAGCCACGTCG	385
QY	818	CCACAGAGCTGAATCTCCACTGCTGTCTCCCTCGTTCGGGTGGCCGCGATGCAGG	877
Db	384	CCCTGGAGAAAGCGCTGCCCGCTGTCTTCCTCCGCTCCGGCGGGCGCGGCATGCAGG	325
QY	878	AAGACATTCAGCTTTTGTGATGATGGTGTGCATTAACCGGGCGTGCAGCGTCACGCG	937
Db	324	AAGGCGTGAATTCCTGATGCAGATGGCCAAAGACTTGCGGGTCTTGCGCGCTCGCGG	265
QY	938	AGGCGCATTTGCCCCTTCTGGTGTATTTTGGCAATCCACGATGGATGGCGCCATGGCT	997
Db	264	AAGAAAGGATCCGCTTGTGCTCGGTATTTGAACGACCCGGCTTACGGCGGCGTTTCCGCA	205
QY	998	CGTGGGGTTCATCTGGGGCATCTACTTTTGGCGAAACCCGGCGCGCAATAGTTTCTCTGG	1057
Db	204	GCTTGGGATGCTCTGGGACCGTGTATCGTCGGCGAAACCAAGGCGCTATATGGCTTTCGCG	145
QY	1058	GTCTCGCGGTGGAGTTAACCATCTGGGCAATGCGGCTTCCAGAGCGTGGCAGCAGGCGG	1117
Db	144	GTCCCCCGGTATGAGACAGACCCTCCGCCGAAGAGCTGCGGGAAGGCTTCCAGGATAGCG	85
QY	1118	AGAAATTTGGTGAATACTGCTGTGATTTGATGGAATTTGTGTGCGCCACTCCAAATGGTGTGAG	1177
Db	84	AGTTCTCTCTTGAACATATGGCGCCATTCACATGATCTGTGACATCTGACGAGTTTGGCGCGC	25
QY	1178	CGGTGGCAAA 1187	
Db	24	GCTGTGGCCA 15	

```

1      RESULT 11
2      US-08-961-527-7/c
3      Sequence 7, Application US/08961527
4      Patent No. 6420135
5      GENERAL INFORMATION:
6      APPLICANT: Charles Kunsch
7      TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequence
8      NUMBER OF SEQUENCES: 391
9      CORRESPONDENCE ADDRESS:
10     ADDRESSEE: Human Genome Sciences, Inc.
11     STREET: 9410 Key West Avenue
12     CITY: Rockville
13     STATE: Maryland
14     COUNTRY: USA
15     ZIP: 20850
16     COMPUTER READABLE FORM:
17     MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
18     COMPUTER: HP Vectra 486/33
19     OPERATING SYSTEM: MSDOS version 6.2
20     SOFTWARE: ASCII Text
21     CURRENT APPLICATION DATA:
22     APPLICATION NUMBER: US/08/961,527
23     FILING DATE:
24     CLASSIFICATION: 424
25     PRIOR APPLICATION DATA:
26     APPLICATION NUMBER:
27     FILING DATE:
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Brookes, A. Anders
30     REGISTRATION NUMBER: 36,373
31     REFERENCE/DOCKET NUMBER: PB340P1
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (301) 309-8504
34     TELEFAX: (301) 309-8512
35     INFORMATION FOR SEQ ID NO: 7:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 19702 base pairs
38     TYPE: nucleic acid
39     STRANDEDNESS: double

```

TOPOLOGY: linear

US-08-961-527-7
Query Match 4.5%; Score 96.2; DB 4; Length 19702;
Best Local Similarity 49.5%; Pred. No. 1.1e-17;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAAATTAATGAGAGGACCGGTGAGGAGCAATCCGCTAGCCGTTATTTTGT 745
DB 11553 ATGAAGCCGTTGACAGGAACTGCTTATTAAGTGCAGACTGCGCTTTGGGATTA 11494
QY 746 CCGATTTTCTCTCTCGCGCGGTTCTTTGGGCAAGGTCGCGTGCATCAGTAAG 805
DB 11493 TGGATTTCTAATTTTATCATGGCTTCTATGGGTACGGTTGAGTGAATAATCATCGTT 11434
QY 806 CGATTCACCGCGCCACAGAGCTGAAATCCCACTGCTGCTCCCTGCTTCCGCTG 865
DB 11433 TGTTCAGTATCGACTGTCGAAATAATGCCAGTTGCTCTATTCACAGCCCTCTGTGGAG 11374
QY 866 CGCGCATGAGAGAGAACATCGAGCTTTTGTATGATGTCATTAACCGGCGCTGTGC 925
DB 11373 CCCGATGAGAGAGAAATCATGAGTCTCATGCAATGCTAAGATCTCTGCGCGGTTA 11314
QY 926 AGCTTCACCGCGAGGCGCATTTGCCGTTCTGTGTATTTGGCAATCCCAAGATGGTG 985
DB 11313 AAGCGCATTCAAATGCTGGTCTCTTTTACCTGACCATTTTGAAGATCCAGCATGGTG 11254
QY 986 GCGCATGCGCTCGTGGGTTTCACTGCGCATCTCACTTTGCGGAAACCGCGCGCAG 1045
DB 11253 GTGACAGACTCTTCTTCTGATGAGAGGAGATATCAATCTGCTGAAACACAGAGCTTG 11194
QY 1046 TAGTTTCTCGGTCCTCGCTGCTGAGTAACTGAGCATCTGCGGATGCGCTTCCAGCGTG 1105
DB 11193 TTGTTTCTGCTGGGCTGCTGATGATTAAGATCGTTGTAAGCTTCCCTGAGGATT 11134
QY 1106 TGCAGCAGCGGAGAAATTTGTAATACTGTGTATGATGAAATTTGTGCCACTCC 1165
DB 11133 TCCAAAGGCGAATTTCTTATTAAGAACATGGCTTTGTGATGCTATTTGCAAAAGAG 11074
QY 1166 AATGCGTGCAGCGGTGCA 1186
DB 11073 ACTTACAGATGATGATGCTA 11053

RESULT 12

US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD

COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB18693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1
Query Match 4.5%; Score 96.2; DB 4; Length 1830121;
Best Local Similarity 49.5%; Pred. No. 1.2e-16;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAAATTAATGAGAGGACCGGTGAGGAGCAATCCGCTAGCCGTTATTTTGT 745
DB 1336980 AAGATGCCCTAATTAATTAATGACAGGTACATTTATATATGCAATCGTTGGCTGCAT 1337039
QY 746 CCGATTTTCTCTCTCGCGCGGTTCTTTGGGCAAGGTCGCGTGCATCAGTAAG 805
DB 1337040 CGAATTTTCTTTTATGAGCGGTTCAATGGCTTCTGTATGTTGTGCAAAATTTGTTAAG 1337099
QY 806 CGATTCACCGCGCCACAGAGCTGAAATCCCACTGCTGATCTCCCTCTTCCGCTG 865
DB 1337100 CGGCTGAAAGGCGATGAAATGATGTCATTTGTGTTTCTCTGAGTGGTG 1337159
QY 866 CGCGATGAGAGAGAACATGAGCTTTTGTATGATGTCATTAACCGCGCTGTGC 925
DB 1337160 CTGATGACAGAGAACATTAATTTCTTTAATGCAAAATGCAAAACTGATGCGTACTG 1337219
QY 926 AGCTTCACCGCGAGGCGCATTTGCCGTTCTGTGTATTTGGCAATCCCAAGATGGTG 985
DB 1337220 CTCAAATGCTGTAAGAGGCTGTGCCATTTATTAATGATTAAGGATCCGATTTAAGGG 1337279
QY 986 GCGCATGCGCTCGTGGGTTTCACTGCGCATCTCACTTTTGGGAAACCGCGCGCAG 1045
DB 1337280 GCGTATCAGCCAGTTTTCGATGATGAGGATTTAATTAATTTGCGAGCAAAAGCTTAA 1337339
QY 1046 TAGTTTCTCGGTCCTCGCTGCTGAGTAACTGAGCATCTGCGGATGCGCTTCCAGCGTG 1105
DB 1337340 TTGTTTTCAGAGGCGCAAGCTTTATTAAGAACAACTGTGCGTGAATAATTTGCCAGAGGTT 1337399
QY 1106 TGCAGCAGCGGAGAAATTTGTAATACTGTGTGATGATGAAATTTGTGCCACTCC 1165
DB 1337400 TCCAACTGATGATGATTTCTTATGAGAAAGGCGCAATGATGATGATGTAAGCGTTGAG 1337459
QY 1166 AATGCGTGCAGCGGTGCA 1186
DB 1337460 AATGCGTGCAGACTTTGCA 1337480

RESULT 13

US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 20:26:35 ; Search time 1275.22 Seconds
(without alignments)
5440.955 Million cell updates/spc

Title: US-10-024-370-1

Perfect score: 2123

Sequence: 1 ctccgagcgsgagtcgctgcatc.....gacctcatgcctggatc 2123

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US05_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	2123	US-10-024-370-1	Sequence 1, Appli
2	2111	99.4	3309400	US-09-738-626-1	Sequence 1, Appli
3	1473	69.4	1473	US-09-738-626-929	Sequence 929, App
4	1473	69.4	1473	US-10-024-370-2	Sequence 2, Appli
5	312.6	14.7	1488	US-09-712-363-35	Sequence 35, Appli
6	131	6.2	354	US-09-738-626-3442	Sequence 3442, Ap
7	122.8	5.8	873	US-09-815-242-7800	Sequence 7800, Ap
8	121	5.7	867	US-09-815-242-6828	Sequence 6828, Ap
9	121	5.7	6021	US-09-070-927A-458	Sequence 458, App
10	119.2	5.6	870	US-09-815-242-7288	Sequence 7288, App
11	119.2	5.6	1053	US-09-895-913A-97	Sequence 97, Appli
12	108	5.1	993	US-09-815-242-9854	Sequence 9854, Ap
13	107.8	4.8	915	US-09-815-242-6133	Sequence 6133, Ap
14	97.8	4.6	867	US-09-815-242-9520	Sequence 9520, Ap
15	96.8	4.6	654	US-09-974-300-5531	Sequence 5531, Ap
16	96.2	4.5	891	US-09-815-242-7115	Sequence 7115, Ap

17	96.2	4.5	1830121	US-10-329-960-1	Sequence 1, Appli
18	93	4.4	867	US-09-815-242-9267	Sequence 9267, Ap
19	86.4	4.1	858	US-09-815-242-8520	Sequence 8520, Ap
20	85.2	4.0	756	US-09-815-242-4369	Sequence 4369, Ap
21	84.8	4.0	7972	US-08-781-986A-312	Sequence 312, App
22	81.8	3.9	951	US-09-815-242-7836	Sequence 7836, Ap
23	79	3.7	1395	US-10-156-761-5702	Sequence 5702, Ap
24	79	3.7	9025608	US-10-156-761-1	Sequence 1, Appli
25	67.8	3.2	1596	US-10-156-761-3318	Sequence 3318, Ap
26	67.8	3.2	9025608	US-10-156-761-1	Sequence 1, Appli
27	59.6	2.8	789	US-09-815-242-6829	Sequence 6829, Ap
28	59.6	2.8	798	US-09-815-242-3309	Sequence 3309, Ap
29	57.2	2.7	971	US-09-974-300-1085	Sequence 1085, Ap
30	56.4	2.7	7989	US-09-070-927A-269	Sequence 269, App
31	54.6	2.6	960	US-09-815-242-5951	Sequence 5951, Ap
32	52.6	2.5	960	US-09-815-242-9799	Sequence 9799, Ap
33	49.8	2.3	768	US-09-815-242-9268	Sequence 9268, Ap
34	49.8	2.3	768	US-09-815-242-9521	Sequence 9521, Ap
35	49.8	2.3	1128	US-10-156-761-3349	Sequence 3349, Ap
36	48.6	2.3	230	US-09-815-242-2856	Sequence 2856, Ap
37	48.6	2.3	1629	US-10-156-761-3330	Sequence 3330, Ap
38	48.2	2.3	984	US-09-767-479-11	Sequence 11, Appli
39	46.2	2.2	618	US-10-003-446-14	Sequence 14, Appli
40	46	2.2	975	US-09-841-132-461	Sequence 461, App
41	45.8	2.2	2325	US-09-815-242-4026	Sequence 4026, App
42	45	2.1	1125	US-10-156-761-2173	Sequence 2173, Ap
43	44.8	2.1	951	US-10-156-761-7101	Sequence 7101, Ap
44	44.4	2.1	78604	US-10-162-497-7	Sequence 7, Appli
45	43.4	2.0	2101	US-10-289-757-151	Sequence 151, App

ALIGNMENTS

RESULT 1
US-10-024-370-1
Sequence 1, Application US/10024370
Publication No. US20020142405A1
GENERAL INFORMATION:
APPLICANT: TILG, YVONNE
APPLICANT: ELKMANN, BERND
APPLICANT: EGGELING, LOTMAR
APPLICANT: SAHM, HERMANN
APPLICANT: MCKEL, BETTINA
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
FILE REFERENCE: 21123-284139-MAS
CURRENT APPLICATION NUMBER: US/10/024,370
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/362,899
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: DE 199 24 365.4
PRIOR FILING DATE: 1999-07-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2123
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: gene
LOCATION: (508)..(1980)
OTHER INFORMATION: accDA
US-10-024-370-1

Query Match 100.0%; Score 2123; DB 13; Length 2123;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTCGAGCGGAGTGGTGATCGGCACCTCTTAAGCATGCGGCTTTAAATAAGCA 60
Db 1 CTCGAGCGGAGTGGTGATCGGCACCTCTTAAGCATGCGGCTTTAAATAAGCA 60

QY 61 CTATATGTTTCTGACCATCTGGCCGACGACCAAGATATGTTGATCAGACTA 120
Db 61 CTATATGTTTCTGACCATCTGGCCGACGACCAAGATATGTTGATCAGACTA 120
QY 121 AACGCTGAATGTGAAGTTACTTAATCTACATTTGCAATGAGCATAGTGAAGATC 180
Db 121 AACGCTGAATGTGAAGTTACTTAATCTACATTTGCAATGAGCATAGTGAAGATC 180
QY 181 ACTCCCCCAATATCTTAACCTTAAAGTATGTTTAACTGATTTATTAAGT 240
Db 181 ACTCCCCCAATATCTTAACCTTAAAGTATGTTTAACTGATTTATTAAGT 240
QY 241 TCCCGATTAACCCCTCTTAAACCCCGAAATACCCCTTTTGCAAGATGCAACACCA 300
Db 241 TCCCGATTAACCCCTCTTAAACCCCGAAATACCCCTTTTGCAAGATGCAACACCA 300
QY 301 GTGCAATAGTTAAAGGCTTCAACGTCACCATCTGTCGGTTTAAAGCTATGTCGG 360
Db 301 GTGCAATAGTTAAAGGCTTCAACGTCACCATCTGTCGGTTTAAAGCTATGTCGG 360
QY 361 ACGTCTAGCAAAAAGTATGTTGTGAATGAAGCAATATCCGTCAATTTTAAAGCA 420
Db 361 ACGTCTAGCAAAAAGTATGTTGTGAATGAAGCAATATCCGTCAATTTTAAAGCA 420
QY 421 TCGATAGCTAAATTTGGGCTTAAATCTTCCGCTTAATAGTATGAGAGACATTGGA 480
Db 421 TCGATAGCTAAATTTGGGCTTAAATCTTCCGCTTAATAGTATGAGAGACATTGGA 480
QY 481 ATTAATTAACAAGCCATTTTTCGCGCGTGAAGAGCTTTTCCGACTATGTTGAGG 540
Db 481 ATTAATTAACAAGCCATTTTTCGCGCGTGAAGAGCTTTTCCGACTATGTTGAGG 540
QY 541 ATGGAACAACCTTCAAGATTGACGCTCAATGACTGCTTGAAGACCTTCAAT 600
Db 541 ATGGAACAACCTTCAAGATTGACGCTCAATGACTGCTTGAAGACCTTCAAT 600
QY 601 TCTTGAATGAAGCTCCCAATATGACCACTCAATCAAGGCTATGAGAGACCTTGA 660
Db 601 TCTTGAATGAAGCTCCCAATATGACCACTCAATCAAGGCTATGAGAGACCTTGA 660
QY 661 CGGCTCGAAGCAAGGCCAAATGCGATGATCGTAACTTGAAGAGGACCGTGA 720
Db 661 CGGCTCGAAGCAAGGCCAAATGCGATGATCGTAACTTGAAGAGGACCGTGA 720
QY 721 GGCATTCGGTAAACCGTTATTTTGTCCGATTTTCTTCCGCGGCTTCTTGGGACG 780
Db 721 GGCATTCGGTAAACCGTTATTTTGTCCGATTTTCTTCCGCGGCTTCTTGGGACG 780
QY 781 GTCCGCTGAGGCGCATGATGAAGGCGATTGACCGGCGCACAGACCTGAATCCCA 840
Db 781 GTCCGCTGAGGCGCATGATGAAGGCGATTGACCGGCGCACAGACCTGAATCCCA 840
QY 841 CTGCTCTCCCTGCTTCCGCTGCTGCGGCGATGACGAGAAACAATCGACTTTTGT 900
Db 841 CTGCTCTCCCTGCTTCCGCTGCTGCGGCGATGACGAGAAACAATCGACTTTTGT 900
QY 901 ATGCTGTCATTAACCGGCGCTGTGACGCTGACCGGAGCGCATTTCCGTTCTG 960
Db 901 ATGCTGTCATTAACCGGCGCTGTGACGCTGACCGGAGCGCATTTCCGTTCTG 960
QY 961 TATTTGGCCAAATCCCAAGATGGGTGGCCCAATGCGCTGTGGGGTTCAATCGG 1020
Db 961 TATTTGGCCAAATCCCAAGATGGGTGGCCCAATGCGCTGTGGGGTTCAATCGG 1020
QY 1021 ACTTTTGGGAAACCGGCGCGAGATAGTATCTTCCGAGGCTTCCGCTGCTGAGT 1080
Db 1021 ACTTTTGGGAAACCGGCGCGAGATAGTATCTTCCGAGGCTTCCGCTGCTGAGT 1080
QY 1081 ACTGGGATGCGCTTCCAGACGCTGTGACGAGGCGGAGAAATTTGGTGAAGCTG 1140
Db 1081 ACTGGGATGCGCTTCCAGACGCTGTGACGAGGCGGAGAAATTTGGTGAAGCTG 1140

QY 1141 ATTGATGAATTTGTTGCGCCACTTCCAAATGCGTGAAGCGGTTGGCAAAACCTCAAG 1200
Db 1141 ATTGATGAATTTGTTGCGCCACTTCCAAATGCGTGAAGCGGTTGGCAAAACCTCAAG 1200
QY 1201 ATTGAGCGGCTGAGAGGACGAGATCGTTTCTTCCAAACAATCTCGCGTGAAGCT 1260
Db 1201 ATTGAGCGGCTGAGAGGACGAGATCGTTTCTTCCAAACAATCTCGCGTGAAGCT 1260
QY 1261 GTGATGAGGCGATTTGCGCTTCTGTCGACCCGACAGGCTTGAATTCGGGAGAT 1320
Db 1261 GTGATGAGGCGATTTGCGCTTCTGTCGACCCGACAGGCTTGAATTCGGGAGAT 1320
QY 1321 GAAACGTTGGGGGAGAGCGTGTCAAGCTTTCGATGCGGCTGCGATGAGGCGG 1380
Db 1321 GAAACGTTGGGGGAGAGCGTGTCAAGCTTTCGATGCGGCTGCGATGAGGCGG 1380
QY 1381 GCTGTGCGCTTGGCTGCGGCGATGCGGAGCGGCGGCTGCTGATGTTGAGCAG 1440
Db 1381 GCTGTGCGCTTGGCTGCGGCGATGCGGAGCGGCGGCTGCTGATGTTGAGCAG 1440
QY 1441 CGCGCTTCAAGCTTGGGCGCGACAGAGCTGCTTTTCCGCTGCTGAGCTTTCG 1500
Db 1441 CGCGCTTCAAGCTTGGGCGCGACAGAGCTGCTTTTCCGCTGCTGAGCTTTCG 1500
QY 1501 CGGAGCTAAACCTGCGGATCGTGTCAATGACACCTCCGCGCGCAATTGTGCG 1560
Db 1501 CGGAGCTAAACCTGCGGATCGTGTCAATGACACCTCCGCGCGCAATTGTGCG 1560
QY 1561 GCGGCTGAGGAGCTCGGACATCGCAAGCTGATTTGCGGACCTTGTCAAGCTTAT 1620
Db 1561 GCGGCTGAGGAGCTCGGACATCGCAAGCTGATTTGCGGACCTTGTCAAGCTTAT 1620
QY 1621 GCTTCCCTCCCAACCTTGTGCTATTTGCTGACAGGCGCTTGGCGGCTGCG 1680
Db 1621 GCTTCCCTCCCAACCTTGTGCTATTTGCTGACAGGCGCTTGGCGGCTGCG 1680
QY 1681 ATGCTGCGCGCGCATCTGATCTAGCGGCGGCAAAACCGTGGCTGCTCGGATTT 1740
Db 1681 ATGCTGCGCGCGCATCTGATCTAGCGGCGGCAAAACCGTGGCTGCTCGGATTT 1740
QY 1741 GAGGCGCTCGGACATCTTTCGCGACCAACCAACCGCGGAAATCATAGAGCG 1800
Db 1741 GAGGCGCTCGGACATCTTTCGCGACCAACCAACCGCGGAAATCATAGAGCG 1800
QY 1801 CAAGCGCTGACGCGGACGCACTTTTAAGCCAAAGGCTTATCGACGGATTCGCG 1860
Db 1801 CAAGCGCTGACGCGGACGCACTTTTAAGCCAAAGGCTTATCGACGGATTCGCG 1860
QY 1861 ACCGAGCACTTGTGTAAGAAATTTCTGCGACATCGCAACGCGCTCTCGAAT 1920
Db 1861 ACCGAGCACTTGTGTAAGAAATTTCTGCGACATCGCAACGCGCTCTCGAAT 1920
QY 1921 AACCAATCGAGAGGCGGAGCGGACAGTCTTCAACGATTTGAGCGGAG 1980
Db 1921 AACCAATCGAGAGGCGGAGCGGACAGTCTTCAACGATTTGAGCGGAG 1980
QY 1981 TAAAGAAATTTATGCGCTGATCAATGATGATGAACACAGGCTGACGAGCT 2040
Db 1981 TAAAGAAATTTATGCGCTGATCAATGATGATGAACACAGGCTGACGAGCT 2040
QY 2041 GTGGCGGAAACCTCAAGGCGGTAAGGAGCTGCGGGAATGCTGACGAGCTTC 2100
Db 2041 GTGGCGGAAACCTCAAGGCGGTAAGGAGCTGCGGGAATGCTGACGAGCTTC 2100
QY 2101 GCCGACCTTCAATGCTGGAATTC 2123
Db 2101 GCCGACCTTCAATGCTGGAATTC 2123

RESULT 2
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 98/37484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patent ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 99.4%; Score 2111; DB 10; Length 3309400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

post deleted

QY 1 CTCGAGCGGAGTCCGTGATTCGGCCACTCTCTAGCAATGCCGCTTTAAATAAGCA 60
DB 881620 CTCGAGCGGAGTCCGTGATTCGGCCACTCTCTAGCAATGCCGCTTTAAATAAGCA 881561
QY 61 CTTATATGTTCTCACCACTCGGCGGACGACGAGATATGTTGTCATGACAGCTA 120
DB 881560 CTTATATGTTCTCACCACTCGGCGGACGACGAGATATGTTGTCATGACAGCTA 881501
QY 121 AACGTGTGATGTAAGTACCTAATCTCAATTCGATGCAATGCGATTTGGAAAATC 180
DB 881500 AACGTGTGATGTAAGTACCTAATCTCAATTCGATGCAATGCGATTTGGAAAATC 881441
QY 181 ACTCCCCCAATATCTTAATCTTAATTAAGTATGTTTAACTGCAATTTAAAGT 240
DB 881440 ACTCCCCCAATATCTTAATCTTAATTAAGTATGTTTAACTGCAATTTAAAGT 881381
QY 241 TCCCGATCTACCCCTCTTACCCGAAATACCCCTTTGGCAAAAGTTGCCAAACACACA 300
DB 881380 TCCCGATCTACCCCTCTTACCCGAAATACCCCTTTGGCAAAAGTTGCCAAACACACA 881321
QY 301 GTGCAATAGTAAACGGGCTTCAACAGTCACATTCGTCCGGTTTAAAGCTATGTTCCGG 360
DB 881320 GTGCAATAGTAAACGGGCTTCAACAGTCACATTCGTCCGGTTTAAAGCTATGTTCCGG 881261
QY 361 ACGTCTAGGCAAAAGTATGTTTGTGAGTAAAGCATTAATCCGTCATTTTAAAGCAA 420
DB 881260 ACGTCTAGGCAAAAGTATGTTTGTGAGTAAAGCATTAATCCGTCATTTTAAAGCAA 881202
QY 421 TCGATAGCTTAATAGGCTTATGATCTTCCGCTCTAATATAGTATGACAGACATTCGA 480
DB 881201 TCGATAGCTTAATAGGCTTATGATCTTCCGCTCTAATATAGTATGACAGACATTCGA 881142
QY 481 ATTAAATTAACAAGCATTTTTCGGCCGTGAGAAAGCGTTTTCGCACTATGTTGGGGCC 540
DB 881141 ATTAAATTAACAAGCATTTTTCGGCCGTGAGAAAGCGTTTTCGCACTATGTTGGGGCC 881082
QY 541 ATGAAACACACTTCAGCATTTGACGTCATATGCTGTTTGTGACCTTGAAGCTTCATT 600
DB 881081 ATGAAACACACTTCAGCATTTGACGTCATATGCTGTTTGTGACCTTGAAGCTTCATT 881022

QY 601 TCTTGAAATGAAGTCCCAATATGACAACTCAATCAAGGCTATGACAGACTTTGAG 660
DB 881021 TCTTGAAATGAAGTCCCAATATGACAACTCAATCAAGGCTATGACAGACTTTGAG 880962
QY 661 CGGGCTCGAAGCAAGGCCAATGCGATGATGATTAATTAATGAGAAAGCAACCGTGAG 720
DB 880961 CGGGCTCGAAGCAAGGCCAATGCGATGATGATTAATTAATGAGAAAGCAACCGTGAG 880902
QY 721 GGCATTTCCGGTACCGCTTATTTTGTCCGATTTTCTCTCTGCGGCTTTTGGCAGC 780
DB 880901 GGCATTTCCGGTACCGCTTATTTTGTCCGATTTTCTCTCTGCGGCTTTTGGCAGC 880842
QY 781 GTCGCGTGGTGCATCATGAAAGGATTCACCGGCGCACAGAGTGAATCCCACTG 840
DB 880841 GTCGCGTGGTGCATCATGAAAGGATTCACCGGCGCACAGAGTGAATCCCACTG 880782
QY 841 CTGGTCTCCCTGCTTCGCGTGTGCGGATGACAGAAAGACATGACCTTTTGTGATG 900
DB 880781 CTGGTCTCCCTGCTTCGCGTGTGCGGATGACAGAAAGACATGACCTTTTGTGATG 880722
QY 901 ATGGTGTCTATACCGCGCTGTGACAGCTCACCGGAGGCGCATTTTCCCTGCTG 960
DB 880721 ATGGTGTCTATACCGCGCTGTGACAGCTCACCGGAGGCGCATTTTCCCTGCTG 880662
QY 961 TATTTGCGCAATCCACGATGAGGTGCGGCGCATGAGCTGTGAGGATTCATCTGGGATCTC 1020
DB 880661 TATTTGCGCAATCCACGATGAGGTGCGGCGCATGAGCTGTGAGGATTCATCTGGGATCTC 880602
QY 1021 ACTTTTGGGAAACCCGCGCGCGAGATGATTTCTCTGAGTCTCTGCGGTGTGAGTTAAC 1080
DB 880601 ACTTTTGGGAAACCCGCGCGCGAGATGATTTCTCTGAGTCTCTGAGGATTCATCTGGGATCTC 880542
QY 1081 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGCGGCGGAGATTTGGTGAATCTGCTG 1140
DB 880541 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGCGGCGGAGATTTGGTGAATCTGCTG 880482
QY 1141 ATTATGGAATTTGTGCGCACTCCAAATTTGCGTGAAGGCGGCAAAACCTCAAGT 1200
DB 880481 ATTATGGAATTTGTGCGCACTCCAAATTTGCGTGAAGGCGGCAAAACCTCAAGT 880422
QY 1201 ATTACCGCGTGAAGGCAACGATGTTTCTCCAAACAATCTCTGCGGTGACATTCG 1260
DB 880421 ATTACCGCGTGAAGGCAACGATGTTTCTCCAAACAATCTCTGCGGTGACATTCG 880362
QY 1261 GTGATGAGGCGATTTGCGGCTTCTGTGACCCGCGAGGCGCTGGAATGCGGAGATTATG 1320
DB 880361 GTGATGAGGCGATTTGCGGCTTCTGTGACCCGCGAGGCGCTGGAATGCGGAGATTATG 880302
QY 1321 GAAAGCTTGGGGGCGAGAGTGTCAAGCTTTCTGAGGCGGCTGCGGCGCATTTGAGCCCG 1380
DB 880301 GAAAGCTTGGGGGCGAGAGTGTCAAGCTTTCTGAGGCGGCTGCGGCGCATTTGAGCCCG 880242
QY 1381 GCTGTGCGGCTTGCCTGCGCGGCGATCGGAGGCGCGGCGGCTGTGATTTGGGAGAT 1440
DB 880241 GCTGTGCGGCTTGCCTGCGCGGCGATCGGAGGCGCGGCGGCTGTGATTTGGGAGAT 880182
QY 1441 GCGCGCTTACGCTTTGCGCGCGAGAGCTGCTTTTGGCGGTGTGCGCATTTTCTGCGCG 1500
DB 880181 GCGCGCTTACGCTTTGCGCGCGAGAGCTGCTTTTGGCGGTGTGCGCATTTTCTGCGCG 880122
QY 1501 GCGGAGCTAAACCTGCGCATGCTGTCATTCGACCACTCCGCGGCGGAATTTGCGGAG 1560
DB 880121 GCGGAGCTAAACCTGCGCATGCTGTCATTCGACCACTCCGCGGCGGAATTTGCGGAG 880062
QY 1561 GCGGCTGAGAGCTTCGAGTGCAGACTGATTTGCGGCGCACTTTTCCAGCTTTATGAC 1620
DB 880061 GCGGCTGAGAGCTTCGAGTGCAGACTGATTTGCGGCGCACTTTTCCAGCTTTATGAC 880002
QY 1621 GCTTCCCTTCCCACTGTTTGGTCTATTATGCTACGAGCGTTTGGCGGTGCGGCTGCGC 1680
DB 880001 GCTTCCCTTCCCACTGTTTGGTCTATTATGCTACGAGCGTTTGGCGGTGCGGCTGCGC 879942
QY 1681 ATGCTGCCGCGGATCTGTGTACGCGGCGCAAAACGCTGTGCTTCCGCAATGCAACA 1740

Db 1141 ATTGCTCAGGGCGTGTGGCGGTGGCGCGCTGGCCATGCTGCCGCGATCTGCTACGCG 1200
Qy 1708 GCCGAAAACGCGTGTGCTGCGCCGATTCGCAACGAGAGGGGCGCTCGCCATCTCTTCGCG 1767
Db 1201 GCCGAAAACGCGTGTGCTGCGCCGATTCGCAACGAGAGGGGCGCTCGCCATCTCTTCGCG 1260
Qy 1768 GACACCAACCAACGCGCGCGGAATCATAGAGCAGCAAGGCGTGCAGCGCGCACTTTTA 1827
Db 1261 GACACCAACCAACGCGCGCGGAATCATAGAGCAGCAAGGCGTGCAGCGCGCACTTTTA 1320
Qy 1828 AGCCAGGCGCTTATGACGGGATCGTCGCCGAACCGAGCACTTGTGAAGAAATCTC 1887
Db 1321 AGCCAGGCGCTTATGACGGGATCGTCGCCGAACCGAGCACTTGTGAAGAAATCTC 1380
Qy 1888 GGCACAAATCAGCAAGCGCCCTCTCCGAATTCGATTAACAATCCGAGAGGGCGGAGCGGAC 1947
Db 1381 GGCACAAATCAGCAAGCGCCCTCTCCGAATTCGATTAACAATCCGAGAGGGCGGAGCGGAC 1440
Qy 1948 AGTCGCTTCAACGATTTGAGCGTTAGCGCAG 1980
Db 1441 AGTCGCTTCAACGATTTGAGCGTTAGCGCAG 1473

RESULT 4

US-10-024-370-2
Sequence 2, Application US/10024370
Publication No. US20020142405A1

GENERAL INFORMATION:

APPLICANT: TILG, YVONNE
APPLICANT: ELKMANNS, BERNHARD
APPLICANT: EGGELING, LOTHAR
APPLICANT: SAHM, HERMANN
APPLICANT: MCKEL, BETTINA
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
FILE REFERENCE: 21123-284139-MAS
CURRENT APPLICATION NUMBER: US/10/024,370
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/362,899
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: DE 199 24 365.4
PRIOR FILING DATE: 1999-07-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1473
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1473)
OTHER INFORMATION: accda
US-10-024-370-2

Query Match 69.4%; Score 1473; DB 13; Length 1473;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 GTGAGAGCGTTTCCGACTATGCTGTGGGCGATGAGAACACACTTCGATTCAGCGCTC 567
Db 1 GTGAGAGCGTTTCCGACTATGCTGTGGGCGATGAGAACACACTTCGATTCAGCGCTC 60
Qy 568 ATGAGCTGCTTTGAGCCCTGACAGCTTCATTTCTTGGATGAAGAACTCCCAATATGAC 627
Db 61 ATGAGCTGCTTTGAGCCCTGACAGCTTCATTTCTTGGATGAAGAACTCCCAATATGAC 120
Qy 628 AACCTCATCAAGGCTATGAGAGACCTTGGAGCGGCTTCGAGAGAGGCCAAATGCCAT 687
Db 121 AACCTCATCAAGGCTATGAGAGACCTTGGAGCGGCTTCGAGAGAGGCCAAATGCCAT 180
Qy 688 GAATGCTAATTAATGAGAGAGGACCGTGGAGGCGATTCGCTAGCCGTTATTTTGTCC 747

Db 181 GAATGCTAATTAATGAGAGAGGACCGTGGAGGCGATTCGCTAGCCGTTATTTTGTCC 240
Qy 748 GATTTTCTCTCTCTGCGGCTTCTTTGGGCAACGCTGCGTGTGCTGCTGCTGCTGCTGCTGCT 807
Db 241 GATTTTCTCTCTCTGCGGCTTCTTTGGGCAACGCTGCGTGTGCTGCTGCTGCTGCTGCTGCT 300
Qy 808 ATTCACCGCGCACAGAGCTGMAACTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867
Db 301 ATTCACCGCGCACAGAGCTGMAACTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qy 868 GCGATGAGAGAGCAATTCAGAGCTTTTGTATGATGATGCTTCATTAACCGCGCTGAG 927
Db 361 GCGATGAGAGAGCAATTCAGAGCTTTTGTATGATGATGCTTCATTAACCGCGCTGAG 420
Qy 928 GCTACCGCGAGCGCATTTTCCGCTTCTGCTGCTATTTGGCGAATCCACGATGCTGCTGCT 987
Db 421 GCTACCGCGAGCGCATTTTCCGCTTCTGCTGCTATTTGGCGAATCCACGATGCTGCTGCT 480
Qy 988 GCGATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1047
Db 481 GCGATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 1048 GATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1107
Db 541 GATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 1108 CAGCAGCGGAGAGATTTGCTGAGAACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 1167
Db 601 CAGCAGCGGAGAGATTTGCTGAGAACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 660
Qy 1168 TTGCGTGAAGCGGTGGGCAAAACCTCAAGGTTATTCAGCGGTGAGGCAACGATGCT 1227
Db 661 TTGCGTGAAGCGGTGGGCAAAACCTCAAGGTTATTCAGCGGTGAGGCAACGATGCT 720
Qy 1228 TTTTCTCCAACTCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1287
Db 721 TTTTCTCCAACTCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Qy 1288 GACCCGCGAGAGCTTGAATGCGGAGATTTTGAAGAGTTGGGGCGAGAGCTGCTCAAG 1347
Db 781 GACCCGCGAGAGCTTGAATGCGGAGATTTTGAAGAGTTGGGGCGAGAGCTGCTGCTCAAG 840
Qy 1348 CTTTCTGCTGCGGTGCTGCGGAGATTTGAGCCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 1407
Db 841 CTTTCTGCTGCGGTGCTGCGGAGATTTGAGCCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 1408 GGGGCGCGCGCGTGTGCTGATTTGGGAGATGCGCGCTTACAGCTTGGGCGCGAGAG 1467
Db 901 GGGGCGCGCGCGTGTGCTGATTTGGGAGATGCGCGCTTACAGCTTGGGCGCGAGAG 960
Qy 1468 CTGCGTTTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
Db 961 CTGCGTTTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Qy 1528 ATTCATCGACCTCGCGGCGCGAATTCGAGAGGCGCTGAGAGCTGCGATGCGAGC 1587
Db 1021 ATTCATCGACCTCGCGGCGCGAATTCGAGAGGCGCTGAGAGCTGCGATGCGAGC 1080
Qy 1588 TCGATTCGCGCACCTTGTCCAACTTATTCAGAGCTTCCCTCCCAACCGCTTCCGCTCAT 1647
Db 1081 TCGATTCGCGCACCTTGTCCAACTTATTCAGAGCTTCCCTCCCAACCGCTTCCGCTCAT 1140
Qy 1648 ATGCTGAGGGCGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1707
Db 1141 ATGCTGAGGGCGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1200
Qy 1708 GCCGAAAACGCGTGTGCTGCGCCGATTCGCAACGAGAGGGGCGCTCGCCATCTCTTCGCG 1767
Db 1201 GCCGAAAACGCGTGTGCTGCGCCGATTCGCAACGAGAGGGGCGCTCGCCATCTCTTCGCG 1260
Qy 1768 GACACCAACCAACGCGCGCGGAATCATAGAGCAGCAAGGCGTGCAGCGCGCACTTTTA 1827

Db 1261 GACACCAACACGCGCGGAATCATPAGAGCAGAGCGGTGACGCGGACGACTTTTA 1320
QY 1828 AGCCAGGCGCTTATGAGCGGATCGTCGCCGAACCGAGCACTTTGTGAGAAATTTCTC 1887
Db 1321 AGCCAGGCGCTTATGAGCGGATCGTCGCCGAACCGAGCACTTTGTGAGAAATTTCTC 1380
QY 1888 GGCACATCAGCAAGCCCTCTCCGAATTTGATACATCCGAGAGGCGGAGCGGAC 1947
Db 1381 GGCACATCAGCAAGCCCTCTCCGAATTTGATACATCCGAGAGGCGGAGCGGAC 1440
QY 1948 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1980
Db 1441 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1473

RESULT 5
US-09-712-363-35
Sequence 35, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 1488
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-35

Query Match 14.7%; Score 312.6; DB 10; Length 1488;
Best Local Similarity 54.0%; Pred. No. 2,1e-92;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

QY 812 ACCGCGCCACAGAGCTGAAACTCCCACTGCTGCTCTCCCTGCTTCCGGTGTGCGGCA 871
Db 272 AGCGGGGACCGCGAGCGGCTGCGGCTACTGCGGTACCAAGCTCGGAGGCAACCGCA 331
QY 872 TGCAAGAAACATCGAGCTTTTGTCAATGATGAGTGCATTAACCGCGGCTGTGAGCGTC 931
Db 332 TGCAAGAAAGCAGCGGTGCGGCTTCTGCAATGTGGAAGATCGTGGCCATCCAGCTGC 391
QY 932 ACCCGAGGCGCATTTTCCGCTTCTGCTGATTTTGGCAATCCCAAGATGGGTGCGCA 991
Db 392 ACAACAGGCGCGCTCCCTACCTGCTATTTTGGCCATCCAGCAACCGGTGAGTTT 451
QY 992 TGCGCTGCTGCGGCTTCAATCGGCGATCTCACTTTTGGGAAACCGGCGGCGAGTAAAGTT 1051
Db 452 TGCGGTGTGCGGCTTCTGCGGCGATCTCACTTTCGCGAGCGGCGGCGCTGATCGCT 511
QY 1052 TCCTGGGTCTCGCGGTGAGTTAACTGAGCGCATGCGCTTCCAGAGCGTGTGAGC 1111
Db 512 TTCTGGGACCAAGGCTCTATGAGTGTCTATGCGCAACCTTCCATCCGGCGTCCAA 571
QY 1112 AGCGGAGAAATTTGTGAAAACTGCTGATGATGATGAAATTTGTGCGCACTCAATTGC 1171
Db 572 CCGCGGAAATCTACGCGCGGATGGAATCATCGACGCGCTGTGCACTGAGCGGCTAC 631
QY 1172 GTGAGCGGTGCGAAAAACCTCAAGTTATTCAGCGCGGTAGAGGCAACGATGTTT 1231
Db 632 GACCGATGCTGATGCTGCTTGAAGGTGCTATGACGCTCCGAAACCGTTTCGCAAC 691
QY 1232 CTCGAACAACCTCTGCGGTGCA---CTTCCGATGATGAGAGCGATTTCCGCTGTG 1288
Db 692 CGCAGAGCGCGCGCGCTGACCGATGTCGCCAAGTGGAGCTGCGTGTGAGCATCGGCG 751
QY 1289 ACCCGAGAGCTTGAATGCGGAGATTTATGAAAGTTTGGGCGAGAGCTGTCAAGC 1348
Db 752 GCGCGAGCGCGCGCGGTGAGCGAGCTATGCGACACGCGCCACGAGCGGCTGTGT 811
QY 1349 TTTGTGCGCGCGTGTGCGGCAATTGAGCCCGGTGCGGCTTCCGCGGCGCATTCG 1408
Db 812 TGTCAAGAACGATCAAGAGGAAAGCG---GGACACGCTGCTGCGCTGCGCTTTG 868
QY 1409 GGGCGCGCGCGTGTGCTGATTTGGGCGAG-----GATGCGCGCTTCAAGC 1453
Db 869 GCGCGCAACCGAGTGTGCTGCGCGAGCAAGAGGAGAGGCGGCGGAGGAAAGCATG 928
QY 1454 TTGGCGCGCAGAGCTGCGTTTTCGCGTGTGCAATTTGCTGCGCGGAGCTTAAAC 1513
Db 929 TCGGCGCGCGCTGCTTACGCGAAGCCGAGCGCGGATGCGCTGCGCGAGCTGTGCC 988
QY 1514 TGCGGATCGGTGATGATGAGACCTCGCGCGCGCAATTGTGCGCAAGCGGCTGAGAGC 1573
Db 989 TGCGGCTGTGCTGTATGACCGCGCGAGCTCGCGCTGTGTGCGCGACCGAAGAG 1048
QY 1574 TCGGATCGCAAGCTGATTTGCGCGCACTTGTTCAGACTTATGACGCTCCCTCCCA 1633
Db 1049 GCGGCGTGGCGCGCAATGCGCATTTGCGCGCGAGCTGTGACGCTGATATCCCGA 1108
QY 1634 CCGTTTGTGCTATTTGTGAGGCGTTGCGGTGCGCGCTGCGCATGCTGCGCGCG 1693
Db 1109 CCGTGTGATCTGTGCGCGCAGGCGAGCGCGCGCGCGCGCGCGCATGTGCGCGCG 1168
QY 1694 ATCTGTCTACGCGCGCGGAAACGCGGTGCTGTCCGATTTGCCACAGAGGCGCGCTCG 1753
Db 1169 ACCGCGTGTGCGCGCACTCCAGCGCTGTGCGCGCTTGTCTCCGAAAGAGCAAGCG 1228
QY 1754 CCATCTCTTTCGCGAGACCAACCGCGCGGAAATCATAGACGCAAGAGCGGTGAGG 1813
Db 1229 CGATGTGTTTCCAGACACTGTCTATGCGCGCAACTGCGCGCGCAAGCATTCGGT 1288
QY 1814 CGACGCACTTTAAGCGCAAGGCTTATGACGAGGATGTGCGCGAAACGAGCATTTG 1873
Db 1289 CCGCGCACTTACTGAAATCGGCGATGTGCAACACTGTGCGGAGTACCCCGAGCGCG 1348

QY 1874 TTGAAGA 1880
DB 1349 CAGACGA 1355

RESULT 6

US-09-738-626-3442/c
Sequence 3442, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OKADA, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 3442
LENGTH: 354
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3442

Query Match 6.2%; Score 131; DB 10; Length 354;
Best Local Similarity 100.0%; Pred. No. 2.1e-32;

Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TGGCTGATCAATGATGATGAACACGAGGTAGCGCCAGACAGTGGGTGCCGAAC 2052
DB 354 TGGCTGATCAATGATGATGAACACGAGGTAGCGCCAGACAGTGGGTGCCGAAC 295
QY 2053 CTCAGGCGCGTAGAGCCTCTGCGGAATGCTCAGCTGACGACGCTCCGACCTTCAT 2112
DB 294 CTCAGGCGCGTAGAGCCTCTGCGGAATGCTCAGCTGACGACGCTCCGACCTTCAT 235
QY 2113 GCGTGGAAATTC 2123
DB 234 GCGTGGAAATTC 224

RESULT 7

US-09-815-242-7800
Sequence 7800, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykend, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7800
LENGTH: 873
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(873)
US-09-815-242-7800

Query Match 5.8%; Score 122.8; DB 9; Length 873;
Best Local Similarity 52.8%; Pred. No. 2.1e-29;

Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAATTAATGAGAGGACACCGGAGGATTCGGTACCGTTATTTGT 745
DB 326 AGGAGCGCTGATGCGATGAGCGGCAAGCTGAGGGATGCGGTGCTGCGCT 385
QY 746 CCGATTTTCTTCTCGCGCGGTTTGTGGGACGCTGCGTGGTGCATGAAAG 805
DB 386 TCGAGTTCTCTTCAATGAGCGGTTGATGAGGCGCATGCGGCGGCTTCTCG 445
QY 806 CGATTCACGCGCCACAGAGGAACTCCACGCTGCTCCCGCTTCCGCTGTG 865
DB 446 CAGCCACGTCGCTTGAAGCGCTGCGCTGATCTGCTTCCGCTTCCGCGG 505
QY 866 CGCGCATCAGAGAGACATGAGCTTTGTATGATGATGATGATGATGATGATG 925
DB 506 CGCGCATCAGAGAGCGGTATGATGATGATGATGATGATGATGATGATGATG 565
QY 926 AGCGTACCGGAGGCGGATTTGCGTTCTGATGATGATGATGATGATGATG 985
DB 566 CGCGCTCGGAGAGAGGATGCGTTGCTGATGATGATGATGATGATGATG 625
QY 986 GCGCATGCGCTCGTGGGTTCTATCTGCGCATCTTCTGCGAAGCGGCGG 1045
DB 626 GCGTTCCGCGACGCTGCGATGCTGCGGAGATGATGATGATGATGATGATG 685
QY 1046 TAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1105
DB 686 TCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 745
QY 1106 TCGACGAGCGGAGAAATTTGTAAGAACTGTGTGATGATGATGATGATG 1165
DB 746 TCGACGAGCGGAGAAATTTGTAAGAACTGTGTGATGATGATGATGATGATG 805
QY 1166 AATTGCGGAGCGGATGCGA 1187
DB 806 AGTTCGCGCGGCTGCGCA 827

RESULT 8

US-09-815-242-6828
Sequence 6828, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.

Qy 925 CAGCGTACCGGAGCGGCGATTCGCGTTCCTGATGATTTGGCGAATCCAGATGGGT 984
Db 5561 CAACGGCATATACAAAGCAGCGTTGGTGTATCTTAGCGTATGACTGATTCAGAGACTGGC 5620
Qy 985 GCGCGCATGCGCTCGTGGGTTTCATCTGGGCATCTCACTTTTGGGAAACCGCGCGGAG 1044
Db 5621 GGTGTACCGGCAAGTTTTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 5680
Qy 1045 ATAGGTTTCCTGGTCTCGTCTCGTCTGATGATGATGATGATGATGATGATGATGATGAT 1104
Db 5661 ATCGGTTTTCCTGGTCTCGTCTCGTCTGATGATGATGATGATGATGATGATGATGATGAT 5740
Qy 1105 GTGACGACGCGGAGATTTGGTGAATACTGGTGTGATGATGATGATGATGATGATGATGAT 1164
Db 5741 TTTCAAGGCGGAGTTTCTTTAGAACATGTTTGTGATGATGATGATGATGATGATGATGAT 5800
Qy 1165 CAATTCGCT 1173
Db 5801 CTTTTCGCT 5809

RESULT 10
US-09-815-242-7288
; Sequence 7288, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7288
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(870)
US-09-815-242-7288

Query Match 5.6%; Score 119.2; DB 9; Length 870;
Best Local Similarity 50.9%; Pred. No. 3.3e-28;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 600 TTCTTGAAATGAATATCCCATATGACAACTCATCAAGCTATGACAGACCTTGA 659
Db 240 TTACGGGCTATATGATCTTTAAATTTGATGATGATGATGATGATGATGATGATGATGAT 299
Qy 660 GCGGCTCGAAGCAAGCCAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719

Db 300 AAATACGAAAAAAGAGATTAACCGCCCAAGCTCATGATGACGGGTGAGGCTAAATCA 359
Qy 720 GGGCATTCGCGTACCGCTTATTTGTCCGATTTTCTTCTCGCGGCTTTTGGGAC 779
Db 360 CCGCATGCGCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
Qy 780 GGTGCGGTTCGTCGCGCATATGAAAGCGATTCACCGGCCCAAGAGCTGAAATCTCCACT 839
Db 420 TGTGAGAGGCGAAAAAGATGTAAGAGCAATGATGATGATGATGATGATGATGATGATGAT 479
Qy 840 GGTGCTTCCCTGCTTCCGCTGTCGCGCATGAGAGCAATGATGATGATGATGATGATGAT 899
Db 480 ATGATGATTTTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
Qy 900 GATGATTCATTAACCGCGCTGTGACAGCTGACCGAGCGCATTTGCGGCTTCTGAT 959
Db 540 AATGCTAAACGAGCGCGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
Qy 960 GATTTTGGCAATCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019
Db 600 GCTTTAAGCGATCCCACTTATGAGGCGCTTACCGCATCTTTTCTTTTAAAGGAGATCT 659
Qy 1020 CACTTTGCGGAACCGCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
Db 660 CATATCCGAGAGCGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719
Qy 1080 CACTGGGATGCGCTTCAGACGCTGTCAGACGCGGAGATTTGTGAAATCTGATG 1139
Db 720 TATAGGGCGGATTTGCTGAGGCGCTTCAACAGCGGAAATTTTATTAAGCATGCTT 779
Qy 1140 GATTGATGAAATGATG 1155
Db 780 GATTGATGATGATG 795

RESULT 11
US-09-895-913A-97
; Sequence 97, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(976)
US-09-895-913A-97

Query Match 5.6%; Score 119.2; DB 10; Length 1053;
Best Local Similarity 50.9%; Pred. No. 3.8e-28;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 600 TTCTTGAAATGAATATCCCATATGACAACTCATCAAGCTATGACAGACCTTGA 659
Db 349 TTACGGGCTATATGATCTTTAAATTTGATGATGATGATGATGATGATGATGATGATGAT 408

```
QY 660 GCGGGCTGCAAGCAAGCCAAATGCGATGAAATCGGTAATTACTGAGAGAGGACCGGTGA 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 AAAATACGAAAAAGAGACTAACCCGCAAGCTCACTGATCAGCGGTGAGGCTAAATCAA 468
QY 720 GGGCAATCCGGTAGCCGTTATTTTTCGATTTTTTCTCCGCGGCTTCCTTTGGGAC 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 CCGCATGCTTTGACATGCTGCTGTTGATTTTATTTAGTTTATGAGGGGAGTTTATGGCTC 528
QY 780 GGTGCGGCTGCTGCGCATCATGAGGCGATTCACCGCCACAGAGCTGAAATCCCACT 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TGTGAGGAGGCAAAAAGATCGTAAGCAATCATGCGCGGTGCTTAAAGAGAGCGTT 588
QY 840 GCTGCTCTCCCTGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 ATTGATTTTTCAGGAGATGAGGGGGGCTAGATGAGTAAGATCCACTTATGCTCATCA 648
QY 900 GATGCTGCTCAATACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AATGCTTAAACGAGCGGCTTTGMAACGATGAGAGGCAAACTCCCTTCATTTTC 708
QY 960 GTATTGCGCAATCCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1019
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GCTTTAAGCAATCCACTTTATGAGGGGCGTTAGCGCATCTTTTGTTTTATGAGGAGTCT 768
QY 1020 CACTTTTTCGGAACCCGCGCGGAGATAGTTTCTGCGGCTGCTGCTGCTGCTGCTGCT 1079
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 CATTATGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 828
QY 1080 CACTGGGAGTCCGCTTCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 888
QY 1140 GATTGATGATTTG 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 GATTGATGATTTG 904
```

```
RESULT 12
US-09-815-242-9854
; Sequence 9854, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9854
; LENGTH: 993
```

```
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (993)
US-09-815-242-9854
```

```
Query Match 5.1%; Score 108; DB 9; Length 993;
Best Local Similarity 49.6%; Pred. No. 1,9e-24;
Matches 276; Conservative 0; Mismatches 280; Indels 0; Gaps 0;
```

```
QY 638 AAGGCTATGAGAGACCTTGTAGAGGAGGCTGAAAGCAAGCCAAATGCGATGAAATCGGTAA 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 AAAAATTAAGACAGACTGCGCTCCGCGAGAAAGAAATCCGCGAGAAAGACGCTGG 403
QY 698 TTACTGAGAGGACCGGTGAGAGGCAATTCGCTGAGCCGTTATTTTGTCCGATTTTCT 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 TGTGATGAAAGGAGCGCTTACGATATGCGGGTTGTCGCGCGGCTTTGAATTCGCGT 463
QY 758 TCCGCGGAGGTTCTTTGGGACAGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 TCATGGGCGCTCAATGGGAGTGTGCTGTTGGCGCAGCTTCGTTCCGCGTTGAACAGG 523
QY 818 CCAAGAGCTGAAACTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 CGCTGAGAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583
QY 878 AAGCAATCGAGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 937
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 AAGGCTGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 643
QY 938 AAGGCTGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 997
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
QY 998 CGTGGGTTTCATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1057
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 704 GTTTTGGATGCTGGGGGATCTCAACATGCGGAGCAAAAGCCGATTTGGCTTGGCCG 763
QY 1058 GTCTCGCGGTGTGAGACTTAACCATGCGGATGCGCTTCCAGACGCTGACAGAGCGG 1117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 GCCCGCGCTTATCGAACAACCGTTGATGAGAGTCCCGCCAGGATTCACGCGAGTG 823
QY 1118 AGAATTGTGAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 AGTTCTGATCGAAGGAGGCTATTGATGATGATGATGATGATGATGATGATGATGATGAT 883
QY 1178 CGGTGGCAAAACCT 1193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 884 AGCTGGGAGACATCT 899
```

```
RESULT 13
US-09-815-242-6133
; Sequence 6133, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 6133
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(915)
US-09-815-242-6133
```

```

Query Match          4.8%; Score 102.8; DB 9; Length 915;
Best Local Similarity 48.9%; Pred. No. 9.9e-23;
Matches 275; Conservative 0; Mismatches 287; Indels 0; Gaps 0;
```

```

QY 707 AAGGACCGTGGAGGCAATCCGGTAGCCGTTATTTTCCGATTTTCTTCCCTCGCG 766
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 335 AAGGACCTCTGATGAAGACCGGTTGTCGCGGCAATTCAGTTCGCTTTAAGGGG 394
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 767 GTTCTTTGGGACGCGTCGCGTGCATCATGAAGCGCATTCACGCGCACAGAGC 826
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 GTTCAATGGGGCTCTTTTGGGTGACAGTTTGTGCGCTTACAGAGGGCTGGAAG 454
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 827 TGAACCTCCACTGCTGTCTCCCTCTTCGCTGCGGTGCGGCGCATGAGAGAACAATC 886
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 455 ATAACCTGCCGCTGATCTGCTTCCGCTCTGCTGCGGCGATGACAGAGACACTGA 514
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 GAGCTTTTGTATGATGATGATCCATACCGCGCTGTGACAGGTACACCGCGGCGCAT 946
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 515 TGTGCTGATGACAGATGCGAAMACCTTGCAGGCACTGCAAMAAATGACAGAGCGGCT 574
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 947 TCCGCTTCCTGATGATTTTGGCGCATCCACAGATGGGTGCGCGCATGCGCTGTGGGTT 1006
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 575 TCCGCTATCATCTCCGCTGCTGACCGACCGAGATGGGGGTTTCTGCAAGTTTCGCA 634
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1007 CATCTGGGCACTCACTTTTGGGAAACCGCGCGCGCAATAGATTCTTGGGCTCTGCG 1066
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 635 TGCTGGGCGATCTCAACATCGCTGAAACCGAAAGCGTTAATCGGCTTTCGCGCTG 694
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1067 TGTGAGACTTAAACCTGGGATGCGCTTCCAGAGGTGTGACAGAGCGGAGAAATTTGG 1126
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 695 TTATCGAAGACAGACGCTTGGCAAAAACGCGCGCTGATTCAGAGCGAGTGAATTCCTGA 754
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1127 TGAACCTGCTGATGATGATGATTTGTGCGCACTCCAAATTTGCGTCAAGCGGAGAA 1186
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 755 TCGAAGAAAGGCGCATGACATGATCTGCTGCTCGGAAATGCGGCTTGAACCTGGGGA 814
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 AAACCTTCAGATTATTCAGCGGAGAGGCAACGATGTTTTTCTCAACAACTCTCG 1246
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 815 GCATTCTGGCGAAGTTGATGATCTGACGCGCGAATCTGAGAGCGCGGTGAAGCG 874
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1247 GCGTGGCACTTCCGCTGATGGA 1268
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 TAGTGTACCCCGGTACCGGA 896
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 14
US-09-815-242-9520
; Sequence 9520, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
```

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykling, Judith W.
; APPLICANT: Mall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 9520
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(867)
US-09-815-242-9520
```

```

Query Match          4.6%; Score 97.8; DB 9; Length 867;
Best Local Similarity 49.7%; Pred. No. 4.4e-21;
Matches 249; Conservative 0; Mismatches 252; Indels 0; Gaps 0;
```

```

QY 686 ATGAATCGGTAATTAAGTGAAGACACCGGAGGCGATTCGGGTAGCCGTTATTTTGT 745
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 ATGAAGCGGTGTGACAGAACTGCTTATTTAAAGTCAACATGTGCTTGGGATTA 391
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 746 CCGATTTTCTTCTCGCGGCTTCTTTTGGACAGGTGCGCTGCGTGCATCATGAAG 805
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 392 TGAATTTCTAATTAATATGAGCTTCTATGAGGTGAGGTGAGGAAATACTCGTT 451
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 806 CGATTACCGGCGCAACAGAGTGAACCTCCACATGCTGTCTCCCTGCTTCCGTTG 865
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 452 TGTTTGAGTATGCACTGTGCAAAAATGCGAGTGTCTTATTCACACCTCTGTGTGAG 511
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 866 CGCGCATGACAGAGAACATGAGCTTTTGTATGATGATGATCCATACCGCGGTGCG 925
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 512 CCGTATGACAGAGAAATATGAGTCTCATGACAGATGCTTAAGTCTTCTGCGGCTTA 571
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 926 AGCGTACCGGAGAGCGGATTTGCGTCTGATGATTTTGGCAATCCACGATGGGTG 985
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 572 AACGTCAATCAAAATGCTGTCTTTTAACTGACATATTTTACAGATCAACGATG 631
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 986 GCGGCAAGCGCTCGTGGGTTCACTGTGCACTCTTTCGGAACCGCGCGGCGAGA 1045
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 632 GTGTGACAGCTTCTTGTGCTATGAGAGCGATATCATCTTGTGCGAACCACAGAGCTTGG 691
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1046 TAGGTTCTGCTGCTGCTGCGTGTGAGATTAACCATGCGGCGATGCGCTTCCAGCGGTG 1105
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 692 TTGCTTTTCTGCGCGCTGCTGTGATGATAAATACGTTTCGGAAGCTTGTGAGATT 751
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1106 TGCAGACGCGGAGAAATTTGTTGAAAACGTGTGATGATGATGAATGTGTGCGCACTCC 1165
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 752 TCCAAAAGGCGAATTCCTATTAACATGAGCTTTGTGAGATGCTATTGTCAAAAAGAG 811
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Oy 1166 ATTGCGTCAGCGGTGCAA 1186
| | | | |
Db 812 ACTTACCGATACGATTGCTA 832

RESULT 15

US-09-974-300-5531
; Sequence 5531, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: fastseq for windows Version 4.0
; SEQ ID NO 5531
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(654)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-5531

Query Match 4.6%; Score 96.8; DB 10; Length 654;
Best Local Similarity 51.8%; Pred. No. 7.8e-21;
Matches 288; Conservative 0; Mismatches 264; Indels 4; Gaps 3;

Oy 636 TCAAGGCTATGAGAGACCTTGGAGCGGCTCGAAGCAAGGCCAATGCGATGATCGT 695
| | | | |
Db 84 TCCAGTATGAG 143
| | | | |
Oy 696 AATTACTGAG 754
| | | | |
Db 144 GGTCACTGATGAG 203
| | | | |
Oy 755 CCTTCTCGGCGGCTCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814
| | | | |
Db 204 ATTTTCGATGAG 263
| | | | |
Oy 815 GGGCCAG 872
| | | | |
Db 264 TGGCCATGAG 323
| | | | |
Oy 873 GCAG 931
| | | | |
Db 324 GCAAG 383
| | | | |
Oy 932 ACCGAG 991
| | | | |
Db 384 TGAACCGAG 443
| | | | |
Oy 992 TGGCTCGTGGGAG 1051
| | | | |
Db 444 CCGCTAGATTGCGGCTCTCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 503
| | | | |
Oy 1052 TCTTGGGTCCTGCGTGTGAGATTACCACTGGGAGAGAGAGAGAGAGAGAGAG 1111
| | | | |
Db 504 TCGCTGGGAGGCGCATTTATGAACAGACCATTCGTGAGAACTTCAAGAGAGAG 563
| | | | |
Oy 1112 AGCGGAG 1171
| | | | |
Db 564 CAGCAGAGATTTCTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 623
| | | | |

Oy 1172 GTGCAGCGGTGCGCAA 1187
| | | | |
Db 624 AAGAAACGCTGACTTA 639

Search completed: November 13, 2003, 06:16:26
Job time : 1310.22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:48:07 ; Search time 5527.03 Seconds
(without alignments)
10902.766 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtgagaagcgtttccgcac.....gattcagcgtttacgcag 1473

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_srs:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1473	100.0	1473	6 AR202320	AR202320 Sequence
2	1473	100.0	1473	6 AX121013	AX121013 Sequence
3	1473	100.0	1473	6 BD004742	BD004742 Replicab1
4	1473	100.0	1473	6 BD163130	BD163130 Novel pol
5	1473	100.0	1575	6 AX066443	AX066443 Sequence
6	1473	100.0	2123	6 CGL17592	Y17592 Corynebacte
7	1473	100.0	2123	6 AR202319	AR202319 Sequence
8	1473	100.0	2123	6 BD004741	BD004741 Replicabl
9	1473	100.0	332050	6 AP005276	AP005276 Corynebac
10	1473	100.0	349980	6 AX127145	AX127145 Sequence
11	602.4	40.9	300750	1 AP005217	AP005217 Corynebac
12	372	25.3	3013	1 CCG1TG	X66112 C.gutamicu
13	312.6	21.2	17810	1 AE006979	AE006979 Mycobacte
14	312.6	21.2	37630	1 MTCY31	B73101 Mycobacteri
15	312.6	21.2	327650	1 BX248337	BX248337 Mycobacte
16	164.4	11.2	342300	1 MLEPRTN8	AL583924 Mycobacte
17	158.4	10.8	10029	1 AE009634	AE009634 Brucella
18	156.8	10.6	10029	1 AE014498	AE014498 Brucella
19	148.6	10.1	260050	1 SME591782	AL591782 Sinorhizo
20	142	9.6	36063	1 MLI15184	U15184 Mycobacteri
21	140.4	9.5	332635	1 AP003005	AP003005 Mesorhizo
22	136.6	9.3	298900	1 AP005937	AP005937 Bradyrhiz
23	132.6	9.0	294250	1 AP001517	AP001517 Bacillus
24	128.4	8.7	7697	1 AE007944	AE007944 Agrobacte
25	128.4	8.7	9508	1 AE008976	AE008976 Agrobacte
26	124.8	8.5	6908	1 SP059237	U59237 Synchococc
27	122.8	8.3	4776	1 AB031231	AB031231 Pseudomon
28	122.8	8.3	11976	1 AE004735	AE004735 Pseudomon
29	122.2	8.3	14401	1 AE001518	AE001518 Helicobac
30	121	8.2	304454	1 AE016956	AE016956 Enterococ
31	120.4	8.2	107440	1 AE012365	AE012365 Xanthomon
32	120.2	8.2	311321	1 NMA322491	AL162754 Neisseria
33	119.2	8.1	1053	6 BD092453	BD092453 Identific
34	119.2	8.1	14160	1 AE000604	AE000604 Helicobac
35	119	8.1	3777	1 AF139661	AF139661 Azospiril
36	118.8	8.1	300975	1 AE017013	AE017013 Bacillus
37	118.8	8.1	301200	1 AP005374	AP005374 Thermosyn
38	118.6	8.1	10377	1 AE002422	AE002422 Neisseria
39	118.6	8.1	349980	6 AX044030	AX044030 Sequence
40	117.2	8.0	291804	1 AE017039	AE017039 Bacillus
41	116.6	7.9	13369	1 AE006013	AE006013 Caulobact
42	113.4	7.7	300363	1 AE016781	AE016781 Pseudomon
43	110.8	7.5	14704	1 AE001970	AE001970 Deinococc
44	110.2	7.5	310967	1 AE016869	AE016869 Pseudomon
45	109.8	7.5	15468	1 AE000690	AE000690 Aquifex a

ALIGNMENTS

RESULT 1
AR202320
LOCUS AR202320 1473 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 2 from patent US 6361986.
ACCESSION AR202320
VERSION AR202320.1 GI:20256859
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.

REFERENCE
1 (bases 1 to 1473)
Tilg, Y., Eikmanns, B., Eggeling, L., Sahn, H. and Mockel, B.
AUTHORS
TITLE
Process for the preparation of L-amino acids by fermentation and
nucleotide sequences coding for the accDA gene
JOURNAL
Patent: US 6361986-A 2 26-MAR-2002;

Pred. No. is the number of results predicted by chance to have a

FEATURES
source
BASE COUNT 275 a 414 c 456 g 328 t
ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred. No. 6,7e-293;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGGAAGCGTTTCCGACTATGTGTGTGGGCGATGGAACACTTCGACATTGACGCTC 60
DB 1 GTGGAAGCGTTTCCGACTATGTGTGTGGGCGATGGAACACTTCGACATTGACGCTC 60
QY 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATATAC 120
DB 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATATAC 120
QY 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGGCCAATGCGAT 180
DB 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGGCCAATGCGAT 180
QY 181 GAATCGTATATTACTGGAAGAGCAACCGTGAAGGCGATTCCGGTATATTTGTCC 240
DB 181 GAATCGTATATTACTGGAAGAGCAACCGTGAAGGCGATTCCGGTATATTTGTCC 240
QY 241 GATTTTCTCTTCTGGGCGGTTCTTTGGGCAAGGTTCGGTTCGGTCCGATATGAAGCG 300
DB 241 GATTTTCTCTTCTGGGCGGTTCTTTGGGCAAGGTTCGGTTCGGTCCGATATGAAGCG 300
QY 301 ATTCAACGGGCGACAGAGCTGAAACTCCCACTGCTGTCTCCCTGCTTCCGGTGTGCG 360
DB 301 ATTCAACGGGCGACAGAGCTGAAACTCCCACTGCTGTCTCCCTGCTTCCGGTGTGCG 360
QY 361 CGCATGACAGAAAGACATGAGCTTTTGTATGATGTGTCCATAACCGCGCTGTGAG 420
DB 361 CGCATGACAGAAAGACATGAGCTTTTGTATGATGTGTCCATAACCGCGCTGTGAG 420
QY 421 CGTCAACCGGAGGCGCATTTGCGTTCCTGTGTATTTGGCGAATCCCAATGGGTGGC 480
DB 421 CGTCAACCGGAGGCGCATTTGCGTTCCTGTGTATTTGGCGAATCCCAATGGGTGGC 480
QY 481 GCCATGAGGCTGTGGGGGTTCACTGAGGCACTCACTTTGGCGAACCAGCGCGAGATA 540
DB 481 GCCATGAGGCTGTGGGGGTTCACTGAGGCACTCACTTTGGCGAACCAGCGCGAGATA 540
QY 541 GGTTCCTGTGTCTCGCGTGTGTGAATTACCACTGGGCGATGCGCTTCAGACGCTGTG 600
DB 541 GGTTCCTGTGTCTCGCGTGTGTGAATTACCACTGGGCGATGCGCTTCAGACGCTGTG 600
QY 601 CAGCAGGCGGAGAAATTTGTGTAAATGTGTGTATGAAATTTGTGTCCCACTCCAA 660
DB 601 CAGCAGGCGGAGAAATTTGTGTAAATGTGTGTATGAAATTTGTGTCCCACTCCAA 660
QY 661 TTGGCGGACGCGGTGGCAAAAACCTCAAGTTATTCAACCGGTGAGGCAAGATGCT 720
DB 661 TTGGCGGACGCGGTGGCAAAAACCTCAAGTTATTCAACCGGTGAGGCAAGATGCT 720
QY 721 TTTTCTCCAACAACCTCTGCGGTGTGCACTTCGGTGTATGAGAGCGCAATTTGCGTCTGT 780
DB 721 TTTTCTCCAACAACCTCTGCGGTGTGCACTTCGGTGTATGAGAGCGCAATTTGCGTCTGT 780
QY 781 GACCCGACAGAGGCTTGGAATCGGGAGATTATGAAAGCTTTGGGGGCAAGCTGTCAG 840
DB 781 GACCCGACAGAGGCTTGGAATCGGGAGATTATGAAAGCTTTGGGGGCAAGCTGTCAG 840
QY 841 CTTTCTGTGCGCGGTGTGCGCATTTGAGCCGCGGTGTGCGGCTTCGCTGAGCGCATC 900
DB 841 CTTTCTGTGCGCGGTGTGCGCATTTGAGCCGCGGTGTGCGGCTTCGCTGAGCGCATC 900
QY 901 GGGGGGCGGCGCGTGTGTGATTTGGGAGAGATCGCGCTTCACGCTTGGGCGGAGAG 960
DB 901 GGGGGGCGGCGCGTGTGTGATTTGGGAGAGATCGCGCTTCACGCTTGGGCGGAGAG 960

```

QY 961 CTGGGTTTGGCGGTGTGCACTTTGCGTGGCCGCGAGCTAAACCTGCGATGCTGTCC 1020
DB 961 CTGGGTTTGGCGGTGTGCACTTTGCGTGGCCGCGAGCTAAACCTGCGATGCTGTCC 1020
QY 1021 ATCATGACACCTCCGCGCGCGAATTGTGCGAGCGGCTGAGAGCTCGGCATGCGAAC 1080
DB 1021 ATCATGACACCTCCGCGCGCGAATTGTGCGAGCGGCTGAGAGCTCGGCATGCGAAC 1080
QY 1081 TCGATTGCGGCACTTGTTCAGCTTATGAGAGCTTCCCTCCCAACCGTTTGGTTCATT 1140
DB 1081 TCGATTGCGGCACTTGTTCAGCTTATGAGAGCTTCCCTCCCAACCGTTTGGTTCATT 1140
QY 1141 ATTGTCAGGCGGTTGGCGGTGTGCGCGCTTGCCATGTGCGCCGATCTGTCTACGCG 1200
DB 1141 ATTGTCAGGCGGTTGGCGGTGTGCGCGCTTGCCATGTGCGCCGATCTGTCTACGCG 1200
QY 1201 GCCGAAAACGCGTGTGCTGCGCATTTGCAACAGAGGCGCTGTGCGCATCTCTTCGCG 1260
DB 1201 GCCGAAAACGCGTGTGCTGCGCATTTGCAACAGAGGCGCTGTGCGCATCTCTTCGCG 1260
QY 1261 GACACCAACACGCGCGGAAATCATAGACGCAAGGCGGTGAGCGGCACTTTTA 1320
DB 1261 GACACCAACACGCGCGGAAATCATAGACGCAAGGCGGTGAGCGGCACTTTTA 1320
QY 1321 AGCCAAAGGCTTATGCAAGGAGATGTGCGGAAACGAGACCTTTGTAAGAAATCTC 1380
DB 1321 AGCCAAAGGCTTATGCAAGGAGATGTGCGGAAACGAGACCTTTGTAAGAAATCTC 1380
QY 1381 GGCAACAATCAGCAACGCTCTTCGAAATTGATTAACAATCCGAGAGGCGGAGCGGAC 1440
DB 1381 GGCAACAATCAGCAACGCTCTTCGAAATTGATTAACAATCCGAGAGGCGGAGCGGAC 1440
QY 1441 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1473
DB 1441 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1473

RESULT 2
AX121013
LOCUS AX121013 1473 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 929 from Patent EP1108790.
ACCESSION AX121013
VERSION AX121013.1 GI:14037728
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum
Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A., Novel polynucleotides Patent: EP 1108790-A 929 20-JUN-2001;
JOURNAL KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
1. 1473
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
BASE COUNT 275 a 414 c 456 g 328 t
ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred. No. 6,7e-293;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGGAAGCGTTTCCGACTATGTGTGTGGGCGATGGAACACTTCGACATTGACGCTC 60
DB 1 GTGGAAGCGTTTCCGACTATGTGTGTGGGCGATGGAACACTTCGACATTGACGCTC 60
QY 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATATAC 120
DB 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATATAC 120

```


Dp	61	ATAGACTCGGTTTGGACCCCTGACAGCTTACTTTCTTGGAAATGAACCTCCCAATATGAC	120
Qy	121	AACCTCAATCAAGGCTATATGACAGACTTTGAGCGGGCTTGAAAGCAAGCCAAATCGAT	180
Dp	121	AACCTCAATCAAGGCTATATGACAGACTTTGAGCGGGCTTGAAAGCAAGCCAAATCGAT	180
Qy	181	GAATGGTAAATTACTGGAAGAAGGCACCGTAGAGGGGATTCGGGTACCCTTAATTTTGTCC	240
Dp	181	GAATGGTAAATTACTGGAAGAAGGCACCGTAGAGGGGATTCGGGTACCCTTAATTTTGTCC	240
Qy	241	GATTTTCTCTTCTCGGCGGTTCTTTGGGACCGGTCCGTCGGTGCATCATGAAGGCG	300
Dp	241	GATTTTCTCTTCTCGGCGGTTCTTTGGGACCGGTCCGTCGGTGCATCATGAAGGCG	300
Qy	301	ATTACCCGGCCACAGAGCTGAATCCCACTGCTGTCTCCCTTCCGTTGCGTGGCG	360
Dp	301	ATTACCCGGCCACAGAGCTGAATCCCACTGCTGTCTCCCTTCCGTTGCGTGGCG	360
Qy	361	CGCATGCGAAGACAAATGAGCTTTTGTCAATGATAGTGTCCATAACCCGCGCTGTGCA	420
Dp	361	CGCATGCGAAGACAAATGAGCTTTTGTCAATGATAGTGTCCATAACCCGCGCTGTGCA	420
Qy	421	CGTCAACCGGAGCCGCAATTTGCCGTTCCGTTGATATTTGCGCAATCCCAAGTGGTGG	480
Dp	421	CGTCAACCGGAGCCGCAATTTGCCGTTCCGTTGATATTTGCGCAATCCCAAGTGGTGG	480
Qy	481	GCCATGACCTCGTGGGGTTCAATCTGGGCACTCACTTTTTCGGAACCCGCGCGCAGATA	540
Dp	481	GCCATGACCTCGTGGGGTTCAATCTGGGCACTCACTTTTTCGGAACCCGCGCGCAGATA	540
Qy	541	GCTTCTCTGCGTCTCTCGTGGTGTGAATTAACCACTGGGCAATGCGCTTCCAGACGGTGTG	600
Dp	541	GCTTCTCTGCGTCTCTCGTGGTGTGAATTAACCACTGGGCAATGCGCTTCCAGACGGTGTG	600
Qy	601	CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATTAAGAAATGTGTGCGCACTCCAA	660
Dp	601	CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATTAAGAAATGTGTGCGCACTCCAA	660
Qy	661	TTGCGTGAGCGGTGGCAAAAAACCTCAAGATTATTCAGCCGGTAGAGGCAACGATCGT	720
Dp	661	TTGCGTGAGCGGTGGCAAAAAACCTCAAGATTATTCAGCCGGTAGAGGCAACGATCGT	720
Qy	721	TTTTCTCCAAACAATCTCTGCGGTGGCACTTCCGGTGTAGTAGAGCGATTGCGGTTCTGT	780
Dp	721	TTTTCTCCAAACAATCTCTGCGGTGGCACTTCCGGTGTAGTAGAGCGATTGCGGTTCTGT	780
Qy	781	GACCCGAGAGGCGTGGAAATCGGGGAGATTATGAAACGTTGGGGGCGAGACTCGTCAG	840
Dp	781	GACCCGAGAGGCGTGGAAATCGGGGAGATTATGAAACGTTGGGGGCGAGACTCGTCAG	840
Qy	841	CTTCTGCTGAGCGCTGTGCGCATTTGAGACCCGCGCTGTGTCGCGTGTGCGCGCATC	900
Dp	841	CTTCTGCTGAGCGCTGTGCGCATTTGAGACCCGCGCTGTGTCGCGTGTGCGCGCATC	900
Qy	901	GGGGGCGGGCCGTGTGTCTGATTGGGCAAGATGCGCCGTTCAAGCTTTGGGCGCAGAG	960
Dp	901	GGGGGCGGGCCGTGTGTCTGATTGGGCAAGATGCGCCGTTCAAGCTTTGGGCGCAGAG	960
Qy	961	CTGGGTTTGGGCGTGTGAGCAATTCGTCGGGCGCGGAGCTTAAACCTTGCCGATCGTGTCC	1020
Dp	961	CTGGGTTTGGGCGTGTGAGCAATTCGTCGGGCGCGGAGCTTAAACCTTGCCGATCGTGTCC	1020
Qy	1021	ATCATGCAACCTCGGCGCGCAATTTGTGCAAGCGGCGTGAAGAGCTCGGCATCGCAAGC	1080
Dp	1021	ATCATGCAACCTCGGCGCGCAATTTGTGCAAGCGGCGTGAAGAGCTCGGCATCGCAAGC	1080
Qy	1081	TCGATTGCGCGACCTTGTCCAAAGTTATCGACGCTCCCTCCACCGTTTCGGTCATT	1140
Dp	1081	TCGATTGCGCGACCTTGTCCAAAGTTATCGACGCTCCCTCCACCGTTTCGGTCATT	1140
Qy	1141	ATTGGTCAAGGCGCTTGGGCGGTGGCGCGCTGGGCAATGCTGCCGATCTGTCTAAGCG	1200
Dp	1141	ATTGGTCAAGGCGCTTGGGCGGTGGCGCGCTGGGCAATGCTGCCGATCTGTCTAAGCG	1200

QY	1201	GCCAAAAACGGTGGCTGTCCGAAATTGCACAAGAAGGCGCTCGGCGCATCTCTTCGC	1260
Db	1201	GCCGAAAACGGTGGCTGTCCGAAATTGCACAAAGGAGCGCTGGCGCATCTCTTCGC	1260
QY	1261	GACACCAACGCGCGGAAATCATAGAGCGACAAAGCGGTGCAGCGCATCTTTTA	1320
Db	1261	GACACCAACGCGCGGAAATCATAGAGCGAACAAGCGGTGCAGCGCATCTTTTA	1320
QY	1321	AGCCAAAGGCTTATCGACGGGATGTGTCGCGAAACCGACCTTGTGTAAGAAATCTC	1380
Db	1321	AGCCAAAGGCTTATCGACGGGATGTGTCGCGAAACCGACCTTGTGTAAGAAATCTC	1380
QY	1381	GGCACAATCAGCAAGCGCTCTCCGAAATTGGATTACAATCCGAGAGGCGCGGACCGCAC	1440
Db	1381	GGCACAATCAGCAAGCGCTCTCCGAAATTGGATTACAATCCGAGAGGCGCGGACCGCAC	1440
QY	1441	AGTGCCTTCAACGATTGAGCGTTTGAAGCGAG	1473
Db	1441	AGTGCCTTCAACGATTGAGCGTTTGAAGCGAG	1473

RESULT 3	BD004742	1473 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	BD004742				
DEFINITION	Replicable DNA, amino acid sequence, <i>Corynebacterium microorganism</i> , shuttle vector, and process for producing L-amino acid.				
ACCESSION	BD004742				
VERSION	BD004742.1	GI:18632703			
KEYWORDS	JP 2001008693-A/2.				
SOURCE	<i>Corynebacterium glutamicum</i>				
ORGANISM	<i>Corynebacterium glutamicum</i>				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1473)	1ike,I., Eggering, L., Rickmans, B., Zamu, H. and Meck, V.	Replacable DNA, amino acid sequence, Corynebacterium microorganism, shuttle vector, and process for producing L-amino acid	Patent: JP 2001008693-A 2 16-04N-2001;	
DEUTSCH HUELS AG, FORSCHUNGSZENTRUM JUELICH GMBH	OS	Corynebacterium glutamicum		

	PD	16-JAN-2001	.		
	Pf	24-MAY-2000 JP 2000I53547			
	PR	27-MAY-1999 DE 19924365.4			
	PI	IVENNE TIJKE, LOTHAR EGGERING, BERNHARD EICKMANS, HERMANN ZAMU,			
	PI	VETTYNA MECKEL.			
	PC	C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, PC			
		C12P13/08			
	PC	C12P13/12, C12P13/20, C12P21/02// (C12N15/09, C12R1:15), (C12N1/21,			
	PC	C12R1:15),			
	PC	(C12P13/04, C12R1:15), (C12P13/06, C12R1:15), (C12P13/08, C12R1:15), PC			
		(C12P13/12, C12R1:15), (C12P13/20, C12R1:15), (C12P21/02, C12R1:15), PC			
		C12N15/00.			
	PC	(C12N15/00, C12R1:15)			
FEATURES					
source					
		. 1. .1473 Location/Qualifiers			
		Location/Qualifiers			
		(1). (.1473).			
	Key				
FH	CDS	/organism="Corynebacterium glutamicum"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:1718"			
BASE COUNT		275 a 414 c 456 g 328 t			
ORIGIN					

Query Match	100.0%;	Score 1473;	DB 6;	Length 1473;
Best Local Similarity	100.0%;	Pred. No. 6.7e-293;		
Matches 1473;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTGGAGAGCGTTTCCGACTATGCTGTGGGGCATGTGAACACACTTCAGCATTGACGCTC	60	

Db	1	GTGGAGAAAGCGTTTCCACATATGATGTGGGGCATGGAACAACATTGACATTAACGCTC	60
Qy	61	ATGACTCGGTTTGGACCCTGACAGCTCTCATTTCTTGGAAATGAACTCCCAATATGAC	120
Db	61	ATGACTCGGTTTGGACCCTGACAGCTCTCATTTCTTGGAAATGAACTCCCAATATGAC	120
Qy	121	AACTCATATCAAGGATATGACAGACCTTGGAGCGGGCTCGAAGCAAGGCCAAATTCGAT	180
Db	121	AACTCATATCAAGGATATGACAGACCTTGGAGCGGGCTCGAAGCAAGGCCAAATTCGAT	180
Qy	181	GAATCGTAATATCTGAGAAAGGCACCGTGGAGGGCATTCGCGTATATTTTGTCC	240
Db	181	GAATCGTAATATCTGAGAAAGGCACCGTGGAGGGCATTCGCGTATATTTTGTCC	240
Qy	241	GATTTTTCCTTCTCGGCGGTTCTTTTGGGCAAGGTGCGGTGCGCATATGAAGCG	300
Db	241	GATTTTTCCTTCTCGGCGGTTCTTTTGGGCAAGGTGCGGTGCGCATATGAAGCG	300
Qy	301	ATTACCCGCGCCACAGAGCTGAAATCCCATCTGCTGTCTCCCTGCTTCCGATGTGCG	360
Db	301	ATTACCCGCGCCACAGAGCTGAAATCCCATCTGCTGTCTCCCTGCTTCCGATGTGCG	360
Qy	361	CGCATCGAGAGAACATCGAGCTTTTGTCAATGATGATGTCATPAACCGCGAGCTGTGAG	420
Db	361	CGCATCGAGAGAACATCGAGCTTTTGTCAATGATGATGTCATPAACCGCGAGCTGTGAG	420
Qy	421	CGTCAACGCGAGGCGCATTTTGGCTTCTGTGTATTTTGGCAATCCCAAGATGGATGCG	480
Db	421	CGTCAACGCGAGGCGCATTTTGGCTTCTGTGTATTTTGGCAATCCCAAGATGGATGCG	480
Qy	481	GCCATGGGCTGTGGGGTTTCATCTTGGGCAATCTCATTTTGGGAAACCCGCGCGCATTA	540
Db	481	GCCATGGGCTGTGGGGTTTCATCTTGGGCAATCTCATTTTGGGAAACCCGCGCGCATTA	540
Qy	541	GATTTCCTGGGGCTCGTGGGTGAGTTTAAACAATGGGCAATGGGCTTCCAGAGGTGTG	600
Db	541	GATTTCCTGGGGCTCGTGGGTGAGTTTAAACAATGGGCAATGGGCTTCCAGAGGTGTG	600
Qy	601	CAGCAGGCGGAGAAATTTGGTGAACCTGGTGTATTTGAATGAAATTTGTGTCCCACTCCA	660
Db	601	CAGCAGGCGGAGAAATTTGGTGAACCTGGTGTATTTGAATGAAATTTGTGTCCCACTCCA	660
Qy	661	TTGCGTGCAGCGGTGGCAAAACCTTCAAGTTATTCAAGCCGGTATGAGGCAACGATGTG	720
Db	661	TTGCGTGCAGCGGTGGCAAAACCTTCAAGTTATTCAAGCCGGTATGAGGCAACGATGTG	720
Qy	721	TTTTTCCAAACAATCTGGGGGTGGCATCTCGGATGATGAGAGGCAATGGCGGTTCTGCT	780
Db	721	TTTTTCCAAACAATCTGGGGGTGGCATCTCGGATGATGAGAGGCAATGGCGGTTCTGCT	780
Qy	781	GACCCGCAAGAGCTGTGAATCGGGAGATTTATGAAAGTTTGGGGCGAGAGTGTGTCAAG	840
Db	781	GACCCGCAAGAGCTGTGAATCGGGAGATTTATGAAAGTTTGGGGCGAGAGTGTGTCAAG	840
Qy	841	CTTTCTGGTGGCGGTGCGCATTTGAGCCCGGCTGTGCGGTTTCCCTGAGCGGCATC	900
Db	841	CTTTCTGGTGGCGGTGCGCATTTGAGCCCGGCTGTGCGGTTTCCCTGAGCGGCATC	900
Qy	901	GGGGGCGCGGCGGTGTGCTGATTTGGGAGAGATCGCCGCTTCAAGCTTTGGGCGGAGAG	960
Db	901	GGGGGCGCGGCGGTGTGCTGATTTGGGAGAGATCGCCGCTTCAAGCTTTGGGCGGAGAG	960
Qy	961	CTGCGCTTTTGGCGGTGTGAGCATTTTGGCTGGGCGCGGAGCTTAAACTGTGCGATGTGTC	1020
Db	961	CTGCGCTTTTGGCGGTGTGAGCATTTTGGCTGGGCGCGGAGCTTAAACTGTGCGATGTGTC	1020
Qy	1021	ATCATCGACACTCCGCGCGCGAATTTGTGCAAGGCGGCTGAAGAGCTTGGCATCGCAAG	1080
Db	1021	ATCATCGACACTCCGCGCGCGAATTTGTGCAAGGCGGCTGAAGAGCTTGGCATCGCAAG	1080
Qy	1081	TGATTTGGCGCGCACTTGTGCCAAGTTATCGACGCTCCCTCCCAACGATTTGGGTACTT	1140
Db	1081	TGATTTGGCGCGCACTTGTGCCAAGTTATCGACGCTCCCTCCCAACGATTTGGGTACTT	1140

QY	1141	ATTGTCAGAGGCCGTGGCCGTGGTGGCGCCTGTGCATCTGCTCCCGCATCTTGTTACGCG	1200
Dd	1141	ATTGTCAGAGGCCGTGGCCGTGGTGGCGCCTGTGCATCTGCTCCCGCATCTTGTTACGCG	1200
QY	1201	GCCAAAAACGGTGGCTGTCCGCAATTGTCCACACAGAAGGGCGCTGTGGCCATCTCTTCGCG	1260
Dd	1201	GCCAAAACGGTGGCTGTCTGTCCGCAATTGTCCACACAGAAGGGCGCTGTGGCCATCTCTTCGCG	1260
QY	1261	GACAACCAACGCGCCGGGAAATCATAGAGCAGACAGGCGTGCAGGGCACGACTTTTA	1320
Dd	1261	GACAACCAACGCGCCGGGAAATCATAGAGCAGACAGGCGTGCAGGGCACGACTTTTA	1320
QY	1321	AGCCAAGGGCTTATCGACGGGATGTGGCCGAACCGAAGCACTTGTGAAGAAATCTC	1380
Dd	1321	AGCCAAGGGCTTATCGACGGGATGTGGCCGAACCGAAGCACTTGTGAAGAAATCTC	1380
QY	1381	GGCAACAATCAGACAAGCCCTCTCCGAATTGGATTAACAATCCGAGAGGGCGGAGCGGAC	1440
Dd	1381	GGCAACAATCAGACAAGCCCTCTCCGAATTGGATTAACAATCCGAGAGGGCGGAGCGGAC	1440
QY	1441	AGTCGCTTACACGATTTGAGCGTTTAGCGCAG	1473
Dd	1441	AGTCGCTTACACGATTTGAGCGTTTAGCGCAG	1473
RESULT 4			
BD161310			
LOCUS	BD161310	1473 bp	DNA
DEFINITION	Novel polynucleotide.		linear PAT 17-JAN-2003
ACCESSION	BD161310.1 GI:27868892		
VERSION	JF 2002191370-A/929.		
KEYWORDS	unidentified		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 1473)		
AUTHORS	Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K., Yokoi,H., Tateishi,N., Senoo,A., Ikeda,M. and Ozaki,A.		
TITLE	Novel polynucleotide		
JOURNAL	Patent: JP 2002191370-A 929 09-JUL-2002;		
COMMENT	KIYOMA HAKKO KOGYO CO LTD OS Corynebacterium glutamicum PN JP 2002191370-A/929 PD 09-JUL-2002 PF 15-DEC-2000 JP 2000405096 PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI, PI KEIKO OCHIAI, PI HARUHIKO YOKOI, NAOKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO PI OZAKI PC C12N15/09, C12N15/09, C07KL4/34, C07KL6/12, C07KL6/40, C12M1/00, FC C12M1/15 PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/ PC 04, C12P13/08, PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC G01N33/566, PC G01N33/569, G01N33/68, G01N37/00//C12P21/08, (C12N1/21, C12R1.15), PC (C12N1/21, C12R1.13), (C12N1/21, C12R1.01), (C12P13/08, C12R1.15), PC C12N15/00, C12N15/00 CC Novel polynucleotide CC Novel polynucleotide FH Key FT source FT Location/Qualifiers 1..1473 Location/Qualifiers 1..1473 /organism='Corynebacterium glutamicum'.		
FEATURES			
source			
BASE COUNT	275 a	414 c	456 g 328 t
ORIGIN			
Query Match		100.0%; Score 1473; DB 6; Length 1473;	

Best Local Similarity 100.0%; Pred. No. 6.7e-293;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTGGAGAGGGCTTTCCGACTATGATGAGGCGCATGGAACAACATTCAGATTAACCTC	60
Db	1	GTGGAGAGGGCTTTCCGACTATGATGAGGCGCATGGAACAACATTCAGATTAACCTC	60
Qy	61	ATAGACTGGTATTTGACCTGACAGCTTATTTCTTGAAATGAATCTCCCAATATGAC	120
Db	61	ATAGACTGGTATTTGACCTGACAGCTTATTTCTTGAAATGAATCTCCCAATATGAC	120
Qy	121	AACCTCAATCAAGGCTATGCAAGACCTTGAGCGGGCTCGAAGCAAGGCCAAATGGAT	180
Db	121	AACCTCAATCAAGGCTATGCAAGACCTTGAGCGGGCTCGAAGCAAGGCCAAATGGAT	180
Qy	181	GAATCGGTAATTAATGAGAGAGCCGTGAGGGGCAATCCGATGACCCGTTATTTGTC	240
Db	181	GAATCGGTAATTAATGAGAGAGCCGTGAGGGGCAATCCGATGACCCGTTATTTGTC	240
Qy	241	GATTTTCTTCTCCGCGGCTTCTTGAGGACGCGTCCGTCGCTGAGGCGCATGGAAGCG	300
Db	241	GATTTTCTTCTCCGCGGCTTCTTGAGGACGCGTCCGTCGCTGAGGCGCATGGAAGCG	300
Qy	301	ATTACCGGCGCCACAAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGTGTGCG	360
Db	301	ATTACCGGCGCCACAAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGTGTGCG	360
Qy	361	CGCATGAGAGAACAAATGAGCTTTTGTATGATGATGATGATGATGATGATGATGATG	420
Db	361	CGCATGAGAGAACAAATGAGCTTTTGTATGATGATGATGATGATGATGATGATGATG	420
Qy	421	CGTCAACCGGAGCGCATTTGCGCTTCTGCTGATTTTGGGCAATCCAGATGGGAGC	480
Db	421	CGTCAACCGGAGCGCATTTGCGCTTCTGCTGATTTTGGGCAATCCAGATGGGAGC	480
Qy	481	GCCATGCGCTCGTGGGCTTCACTGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
Db	481	GCCATGCGCTCGTGGGCTTCACTGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
Qy	541	GGTTTCCTGGGCTCGTGGGCTTCACTGCGCATCTGCTGCTGCTGCTGCTGCTGCTG	600
Db	541	GGTTTCCTGGGCTCGTGGGCTTCACTGCGCATCTGCTGCTGCTGCTGCTGCTGCTG	600
Qy	601	CAGCAGCGCGAGAAATTTGATGAATACTGATGATGATGATGATGATGATGATGATG	660
Db	601	CAGCAGCGCGAGAAATTTGATGAATACTGATGATGATGATGATGATGATGATGATG	660
Qy	661	TTGCGTGCAGCGGTGCAAAAACCTCAAGGTTATTCAGCCGCTAGAGGCAACGATCGT	720
Db	661	TTGCGTGCAGCGGTGCAAAAACCTCAAGGTTATTCAGCCGCTAGAGGCAACGATCGT	720
Qy	721	TTTTTCCCAACAATCTCGGCGTGGCACTTCGCGTATGAGAGCCATTCGCGTTTCGT	780
Db	721	TTTTTCCCAACAATCTCGGCGTGGCACTTCGCGTATGAGAGCCATTCGCGTTTCGT	780
Qy	781	GACCCGAGAGGCTGGAATCGGGAGATTTATGAAACGTTGGGGGAGAGAGCTGCTCAAG	840
Db	781	GACCCGAGAGGCTGGAATCGGGAGATTTATGAAACGTTGGGGGAGAGAGCTGCTCAAG	840
Qy	841	CTTTTGGTGGCGGTGCTGCGCATTTAGCCCGCTGTCGCGCTGCTGCGCGCATC	900
Db	841	CTTTTGGTGGCGGTGCTGCGCATTTAGCCCGCTGTCGCGCTGCTGCGCGCATC	900
Qy	901	GGGGGCGGGCGCGTGTGCTGATTTGGGAGAGATTCGCGCTTACGCTTGGGGCGAGAG	960
Db	901	GGGGGCGGGCGCGTGTGCTGATTTGGGAGAGATTCGCGCTTACGCTTGGGGCGAGAG	960
Qy	961	CTGCGTTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020
Db	961	CTGCGTTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020
Qy	1021	ATCATGACACTTCGGGCGCGCAATTTGTCAGGCGGCTGAGAGCTCGGCAATCGCAAC	1080
Db	1021	ATCATGACACTTCGGGCGCGCAATTTGTCAGGCGGCTGAGAGCTCGGCAATCGCAAC	1080

Db	1021	ATCATGACACTTCGGGCGCGCAATTTGTCAGGCGGCTGAGAGCTCGGCAATCGCAAC	1080
Qy	1081	TCGATTGCGGCGACCTTTGTCAAGCTTATGAGGCTCCCTCCACCGTTTGGTCAAT	1140
Db	1081	TCGATTGCGGCGACCTTTGTCAAGCTTATGAGGCTCCCTCCACCGTTTGGTCAAT	1140
Qy	1141	ATTGTCAGGCGCTTGGCGGTGCGGCTGCGCATGCTGCGCGCATCTGCTGCTGCTG	1200
Db	1141	ATTGTCAGGCGCTTGGCGGTGCGGCTGCGCATGCTGCGCGCATCTGCTGCTGCTG	1200
Qy	1201	GCCGAAACCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1260
Db	1201	GCCGAAACCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1260
Qy	1261	GACACCAACGCGCGCGGCAATCATAGAGCGCAAGGCTGAGGCGGCGCATCTTTA	1320
Db	1261	GACACCAACGCGCGCGGCAATCATAGAGCGCAAGGCTGAGGCGGCGCATCTTTA	1320
Qy	1321	AGCCAGGCGCTTATGAGAGGAGATCGTCCGCAACCGAGCATTTGTTGAAGAAATTC	1380
Db	1321	AGCCAGGCGCTTATGAGAGGAGATCGTCCGCAACCGAGCATTTGTTGAAGAAATTC	1380
Qy	1381	GGCACAATGAGCAACGCGCTTCTCCGAATTGATTAACATCCGAGAGGCGGAGCGGAC	1440
Db	1381	GGCACAATGAGCAACGCGCTTCTCCGAATTGATTAACATCCGAGAGGCGGAGCGGAC	1440
Qy	1441	AGTCGCTTCAACGATTTAGCGTTTACGCGAG	1473
Db	1441	AGTCGCTTCAACGATTTAGCGTTTACGCGAG	1473

RESULT 5	AX066443	1575 bp	DNA	linear	PAT 24-JAN-2001
LOCUS	AX066443				
DEFINITION	Sequence 25 from Patent WO0100805.				
ACCESSION	AX066443				
VERSION	AX066443.1	GI:12544151			
KEYWORDS					
SOURCE					
ORGANISM	Corynebacterium glutamicum				
	Corynebacterium glutamicum				
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
	Corynebacterineae; Corynebacteriaceae; Corynebacterium.				
REFERENCE					
AUTHORS	1 Pompeius, M., Kroeger, B., Schroeder, H., Zelder, O. and Habermeyer, G.				
TITLE	corynebacterium glutamicum genes encoding proteins involved in				
JOURNAL	membrane synthesis and transport				
	Patent: WO 0100805-A 25 04-JAN-2001;				
	Basf AKTIENGESSELLSCHAFT (DE)				
FEATURES					
source	Location/Qualifiers				
	1..1575				
	/organism="Corynebacterium glutamicum"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1718"				
	101..1555				
	/note="RXA02173"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CA26449.1"				
	/db_xref="GI:12544152"				
	/translation="MVMGMEHTSALTLIDSVLPDSFISWNETPOYDNLONGYAEITL				
	BARSAKDESVITGEIVGEPVAVILVSDSGSGISGIVASVPIMKAIHRAETIKL				
	PLVSPASGARMEDNRAFYMMVMSITAAVORHREAHPLVLYLRNPTMGAMASWGS				
	SGHLTPAEPMOIGLGRVVELTGHALPGVQAEMLVKTGVIDGIVSPQLASAAV				
	AKTLIVQIVPVAIORFSPITTEGVALPVWEALARSDDPQRPQIGIMETLGDVVLG				
	ARAAGLSPAVVALARIGRPPVILIGDPRFTLPQELRPARGISLARINLPVSI				
	IDTGAELSOAAEELIGASIPARTLSKILIDAPLPVSVITGOGVGGALAMPDLVY				
	AAENWLSALPPEGASALIFPDTHAAELIIRDOGVQAAHLLSGLIDIVAEHFIVE				
	EILITISMAISELNNPERARDRPFRFETLAQ"				
BASE COUNT	309 a	434 c	474 g	358 t	
ORIGIN					
Query Match	100.0%;	Score 1473;	DB 6;	Length 1575;	
Best Local Similarity	100.0%;	Pred. No. 6.6e-293;			

Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	GTGGAAGCGTTTCCGACTATGTTGGGGGCAATGAAACACACTTCACGATTGACGCTC	60	
Dd	80	GTGGAAGCGTTTCCGACTATGTTGGGGGCAATGAAACACACTTCACGATTGACGCTC	139	
Qy	61	ATAGACTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAATGAAATCCCAATATGAC	120	
Dd	140	ATAGACTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAATGAAATCCCAATATGAC	199	
Qy	121	AACCTCAATCAAGGCTATGACAGACCTTGAAGCGGGCTCGAAGCAAGGCCAATGCGAT	180	
Dd	200	AACCTCAATCAAGGCTATGACAGACCTTGAAGCGGGCTCGAAGCAAGGCCAATGCGAT	259	
Qy	181	GAATGGTAATTAATCTGAGAGAGGCACTGAGAGGGCAATTCGGTACCGCTTATTTTGTCC	240	
Dd	260	GAATGGTAATTAATCTGAGAGAGGCACTGAGAGGGCAATTCGGTACCGCTTATTTTGTCC	319	
Qy	241	GATTTTCTCTTCCTGGGGGCTTCTTTGGGCAAGTGGGCTGGGTCGATCATGAAGGCG	300	
Dd	320	GATTTTCTCTTCCTGGGGGCTTCTTTGGGCAAGTGGGCTGGGTCGATCATGAAGGCG	379	
Qy	301	ATTCAACCGGCGACAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCGGTGGTGGC	360	
Dd	380	ATTCAACCGGCGACAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCGGTGGTGGC	439	
Qy	361	CGCATGCAAGAAACATGAGCTTTTGTATGATGGTGTCCATTAACCGGCTGTGAG	420	
Dd	440	CGCATGCAAGAAACATGAGCTTTTGTATGATGGTGTCCATTAACCGGCTGTGAG	499	
Qy	421	CGTCAACCGGCGAGGCGCATTTTGGCGTTTCCGTGGTGTATTTGGCGCAATCCCAAGTGGGTGAC	480	
Dd	500	CGTCAACCGGCGAGGCGCATTTTGGCGTTTCCGTGGTGTATTTGGCGCAATCCCAAGTGGGTGAC	559	
Qy	481	GCCATGAGCTCTGGGGTTCATCTGAGGCACTCACTTTTGGGAAACCGGCGCGCAGATA	540	
Dd	560	GCCATGAGCTCTGGGGTTCATCTGAGGCACTCACTTTTGGGAAACCGGCGCGCAGATA	619	
Qy	541	GGTTTCTGGGTCTCCGCTGGTGTGAATTAACACTGGGCAATCCGCTTCCAGACGGGTGG	600	
Dd	620	GGTTTCTGGGTCTCCGCTGGTGTGAATTAACACTGGGCAATCCGCTTCCAGACGGGTGG	679	
Qy	601	CAGCAGCGGAGAAATTTGGTGAACCTGGTGTATGATGAAATTTGTGCGCACTCCAA	660	
Dd	680	CAGCAGCGGAGAAATTTGGTGAACCTGGTGTATGATGAAATTTGTGCGCACTCCAA	739	
Qy	661	TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTCAGCCGGTAGAGGCAACGGATCGT	720	
Dd	740	TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTCAGCCGGTAGAGGCAACGGATCGT	799	
Qy	721	TTTTTCTCCAACTCTCTGGGCTGGCACTTCGGTGTATGAGGCGATTTGCGGTTCTCGT	780	
Dd	800	TTTTTCTCCAACTCTCTGGGCTGGCACTTCGGTGTATGAGGCGATTTGCGGTTCTCGT	859	
Qy	781	GACCCGAGAGGCTGGAATCGGAGAGATTATGAAACGTTGGGGGAGACGTCGTCAAG	840	
Dd	860	GACCCGAGAGGCTGGAATCGGAGAGATTATGAAACGTTGGGGGAGACGTCGTCAAG	919	
Qy	841	CTTTCGTGCTGGCTGTGCGCATTTGAGCCCGGCTGTGCGGCTTCCCTGGCGGCAATC	900	
Dd	920	CTTTCGTGCTGGCTGTGCGCATTTGAGCCCGGCTGTGCGGCTTCCCTGGCGGCAATC	979	
Qy	901	GGGGGCGGGCGCTGTGCTGTATTTGGGAGAGATCGCCGCTTACGCTTGGGCGGAGAG	960	
Dd	980	GGGGGCGGGCGCTGTGCTGTATTTGGGAGAGATCGCCGCTTACGCTTGGGCGGAGAG	1039	
Qy	961	CTGCGTTTTCGCGCTGTGCGCATTTTCGCGGCGGCGAGCTTAACCTGCGCATGCTGTCC	1020	
Dd	1040	CTGCGTTTTCGCGCTGTGCGCATTTTCGCGGCGGCGAGCTTAACCTGCGCATGCTGTCC	1099	
Qy	1021	ATCATGCAACCTCCGGGCGCCGAATTTGTGCAAGGCGGCTGAGAGAGCTCGGCAAGC	1080	
Dd	1100	ATCATGCAACCTCCGGGCGCCGAATTTGTGCAAGGCGGCTGAGAGAGCTCGGCAAGC	1159	

Qy	1081	TCGATTCGCGGCACTTGTCCAAAGCTTATGACGCTCCCTCCCAAGCTTTGGGTCATT	1140	
Dd	1160	TCGATTCGCGGCACTTGTCCAAAGCTTATGACGCTCCCTCCCAAGCTTTGGGTCATT	1219	
Qy	1141	ATTGGTCAAGGCGCTTGGCGGCTGGCGCGCTGCGCATGCTGCGCGCATGCTTACGCG	1200	
Dd	1220	ATTGGTCAAGGCGCTTGGCGGCTGGCGCGCTGCGCATGCTGCGCGCATGCTTACGCG	1279	
Qy	1201	GCCGAAAACGCTGCTGTCCGCAATTCGACAGAGGGCGCTTGGCCATCTTTCCGC	1260	
Dd	1280	GCCGAAAACGCTGCTGTCCGCAATTCGACAGAGGGCGCTTGGCCATCTTTCCGC	1339	
Qy	1261	GACACCAACCAACGCGCGGGAATATGAGAGGCAAGGCGGCAAGGCGCATTTTA	1320	
Dd	1340	GACACCAACCAACGCGCGGGAATATGAGAGGCAAGGCGGCAAGGCGCATTTTA	1399	
Qy	1321	AGCCAAAGGCTTATCGACGCGGATCGTCCGAAACGAGACATTTGTAAGAAATCTC	1380	
Dd	1400	AGCCAAAGGCTTATCGACGCGGATCGTCCGAAACGAGACATTTGTAAGAAATCTC	1459	
Qy	1381	GGCAATTCAGCAACGCCCTCTCCGAATTGGATTAACATCCGAGAGGGCGGACGCGAC	1440	
Dd	1460	GGCAATTCAGCAACGCCCTCTCCGAATTGGATTAACATCCGAGAGGGCGGACGCGAC	1519	
Qy	1441	AGTGGCTTCAACAGATTGAGCGCTTTAGCGGAG	1473	
Dd	1520	AGTGGCTTCAACAGATTGAGCGCTTTAGCGGAG	1552	

RESULT 6	
CG117592	2123 bp DNA linear BCT 26-JUN-2001
LOCUS	CG117592 2123 bp DNA linear BCT 26-JUN-2001
DEFINITION	Corynebacterium glutamicum accda gene.
ACCESSION	Y17592
VERSION	Y17592.1 GI:14572581
KEYWORDS	accda gene; acetyl-CoA carboxylase; carboxyltransferase.
SOURCE	Corynebacterium glutamicum
ORGANISM	Corynebacterium glutamicum
REFERENCE	1 Kim, K., Peters-Wendisch, P.G., Sahm, H. and Eikmanns, B.J. Unpublished
AUTHORS	Kim, K., Peters-Wendisch, P.G., Sahm, H. and Eikmanns, B.J.
JOURNAL	2 (bases 1 to 2123)
REFERENCE	Peters-Wendisch, P.G.
AUTHORS	Direct Submission
TITLE	Submitted (17-JUN-1998) P.G. Peters-Wendisch, University of
JOURNAL	California at Berkeley, Dept. Plant and Microbial Biology, Kustu
	Lab, 111 Koshland Hall, Berkeley CA 94720, USA
COMMENT	Related sequence x66112.
FEATURES	Location/Qualifiers
source	1..2123
	/organism="Corynebacterium glutamicum"
	/mol_type="genomic DNA"
	/strain="ATCC 13032"
	/db_xref="taxon:1718"
gene	508..1983
	/gene="accda"
CDS	508..1983
	/gene="accda"
	/codon_start=1
	/transl_table=1
	/product="putative carboxyltransferase subunit of
	acetyl-CoA carboxylase"
	/protein_id="CAC42827.1"
	/db_xref="GI:14572582"
	/translation="MEKRPFTMWGMEHTSALTILDSVLDPDSISMENETPOYDNLNO
	GYATTERASAKAKDSEVITGEGTYGIVAVIVATILSPFLGSLGIVASVRIMKAH
	RATELKLPLVSPASGARGQEDNRARVWVMSITAAVQRREAHLPVLVLRNPTWG
	AMASWSSGHLTFABEGAOIQLGAPRVVETLTHALPDQVQAEHLVKTVDIGIYSP
	LOLRAAVKTLKVLQPEVATDRFSPPTGVALLPMEALIASRDPORFGIEIMETLGA
	DVVLTSARAGALSPAVRAVLARIGRPVVLIGDRRFTLGPDLRPARRGISLAREL

NLPVSIIDTSGAELISQAELIASSIARTLSKLIADLPVTVSIIIGGVGAGALAM
 LPADIVYAENMWSALPREGASAILFRDTNHAETIEOGVOAHALLSQGLIDGIVA
 ETEHVEBILIGTISNALSIEDNNPERAGDSRTRERBLAQ"

BASE COUNT 460 a 574 c 592 g 497 t
 ORIGIN

Query Match 100.0%; Score 1473; DB 1; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 6.4e-293;
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGAGAACGCTTTTCCGACTATGTGTGGGCGCATGAAACACCTTCAGCATTTGACGCTC 60
 DB GTGGAGAACGCTTTTCCGACTATGTGTGGGCGCATGAAACACCTTCAGCATTTGACGCTC 567
 61 ATAGACTCGGTTTGGAGCCCTGACGCTTCACTTTCTTGGAACTGAACTCCCAATATGAC 120
 DB ATAGACTCGGTTTGGAGCCCTGACGCTTCACTTTCTTGGAACTGAACTCCCAATATGAC 627
 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGGCCAATGCGAT 180
 DB AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGGCCAATGCGAT 687
 181 GAATGGTATTTACTGAGAAAGGCACTGTGAGAGGCGCATTCGGTAGCCGTTATTTTGTCC 240
 DB GAATGGTATTTACTGAGAAAGGCACTGTGAGAGGCGCATTCGGTAGCCGTTATTTTGTCC 747
 241 GATTTTCTCTCGTGGCGGTTCTTTGGGCACGGTGCGTGGTGGCATTCATGAAAGGCG 300
 DB GATTTTCTCTCGTGGCGGTTCTTTGGGCACGGTGCGTGGTGGCATTCATGAAAGGCG 807
 301 ATTCAACCGGCGCACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGGTGTGCG 360
 DB ATTCAACCGGCGCACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGGTGTGCG 867
 361 CGCATGCGAAGAACATGAGCTTTTGTGATGATGCTGTCCATTAACCGCGGCTGTGCG 420
 DB CGCATGCGAAGAACATGAGCTTTTGTGATGATGCTGTCCATTAACCGCGGCTGTGCG 927
 421 CGTACCGGCGAGCGCATTTGCGCTTCCGTGTATTTGCGGATTCGCAATCCGATGCTGCG 480
 DB CGTACCGGCGAGCGCATTTGCGCTTCCGTGTATTTGCGGATTCGCAATCCGATGCTGCG 987
 928 CGTACCGGCGAGCGCATTTGCGCTTCCGTGTATTTGCGGATTCGCAATCCGATGCTGCG 987
 481 GCCATGCGCTCGTGGGCTTCATCTGGGCACTCACTTTTGGCGAAACCGCGCGCGAGATA 540
 DB GCCATGCGCTCGTGGGCTTCATCTGGGCACTCACTTTTGGCGAAACCGCGCGCGAGATA 1047
 988 GCCATGCGCTCGTGGGCTTCATCTGGGCACTCACTTTTGGCGAAACCGCGCGCGAGATA 1047
 541 GCTTCTCTGCTGCTCGGCTGTGAGATTAACTGAGCATGCGCTTCCAGACGCTGTG 600
 DB GCTTCTCTGCTGCTCGGCTGTGAGATTAACTGAGCATGCGCTTCCAGACGCTGTG 1107
 1048 GCTTCTCTGCTGCTCGGCTGTGAGATTAACTGAGCATGCGCTTCCAGACGCTGTG 1107
 601 CAGCAGCGGCGAATTTGCTGAAACTGTGTGATGATGAAATTGTGTGCGCACTCCAA 660
 DB CAGCAGCGGCGAATTTGCTGAAACTGTGTGATGATGAAATTGTGTGCGCACTCCAA 1167
 1108 CAGCAGCGGCGAATTTGCTGAAACTGTGTGATGATGAAATTGTGTGCGCACTCCAA 1167
 661 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTTATTCAGCGGTAAGAGCAAGATGCT 720
 DB TTGCGTGAAGCGGTGGCAAAACCTCAAGTTTATTCAGCGGTAAGAGCAAGATGCT 1227
 1168 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTTATTCAGCGGTAAGAGCAAGATGCT 1227
 721 TTTTCTCCAACAACCTCTGCGGTGCACTTCCGTTGATGAGCGCATTTGCGGTTCTGCT 780
 DB TTTTCTCCAACAACCTCTGCGGTGCACTTCCGTTGATGAGCGCATTTGCGGTTCTGCT 1287
 1228 TTTTCTCCAACAACCTCTGCGGTGCACTTCCGTTGATGAGCGCATTTGCGGTTCTGCT 1287
 781 GACCGCGAGAGGCTGGGAATTCGGGAGATTAAGAAACGTTGGGGCGAGACTCGTCCAG 840
 DB GACCGCGAGAGGCTGGGAATTCGGGAGATTAAGAAACGTTGGGGCGAGACTCGTCCAG 1347
 1288 GACCGCGAGAGGCTGGGAATTCGGGAGATTAAGAAACGTTGGGGCGAGACTCGTCCAG 1347
 841 CTTTCTGGTGCGGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 900
 DB CTTTCTGGTGCGGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 1407
 1348 CTTTCTGGTGCGGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 1407
 901 GGGGGCGCGCGCTGTGTGATGATGGGCGAGATCGCGCTTCACTGTTGGCGCGAGAG 960
 DB GGGGGCGCGCGCTGTGTGATGATGGGCGAGATCGCGCTTCACTGTTGGCGCGAGAG 1467

QY 961 CTGCGTTTGGCGGTGCGGCACTTTGCTGAGGCGCGAGCTAAACCTGCGCATCGGTCC 1020
 DB 1468 CTGCGTTTGGCGGTGCGGCACTTTGCTGAGGCGCGAGCTAAACCTGCGCATCGGTCC 1527
 QY 1021 ATCATGACACCTCCGCGCGCAATTTGTGAGGCGGCTGAGAGCTCGGCATCGCAAC 1080
 DB 1528 ATCATGACACCTCCGCGCGCAATTTGTGAGGCGGCTGAGAGAGCTCGGCATCGCAAC 1587
 QY 1081 TCGATTGCGCGCACTTTGTCAAGCTTTATGACGCTCCCTTCCCAACGCTTGTGAT 1140
 DB 1588 TCGATTGCGCGCACTTTGTCAAGCTTTATGACGCTCCCTTCCCAACGCTTGTGAT 1647
 QY 1141 ATTGTGAGGCGGTTGGCGGTGCGCGCGCGGCAATGCTCCCGCGCATGCTGATGAC 1200
 DB 1648 ATTGTGAGGCGGTTGGCGGTGCGCGCGCGGCAATGCTCCCGCGCATGCTGATGAC 1707
 QY 1201 GCCGAAACGCGTGTGCTGCTCCGATTTGCAACGAGAGGCGCTCGGCATCTCTTCCGC 1260
 DB 1708 GCCGAAACGCGTGTGCTGCTCCGATTTGCAACGAGAGGCGCTCGGCATCTCTTCCGC 1767
 QY 1261 GACACCAACACGCGCGGAAATCATTAAGCAAGGCGTGCAGGCGCACGCACTTTTA 1320
 DB 1768 GACACCAACACGCGCGGAAATCATTAAGCAAGGCGTGCAGGCGCACGCACTTTTA 1827
 QY 1321 AGCCAAAGGCTTATGAGAGGAGTGTGCTCCGAAACCGAGCACTTTTGAAGAAATTC 1380
 DB 1828 AGCCAAAGGCTTATGAGAGGAGTGTGCTCCGAAACCGAGCACTTTTGAAGAAATTC 1887
 QY 1381 GGCACAAATGAGCAAGCGCTTCCGAATTTGATAAATCCGAGAGGCGGAGCGGAC 1440
 DB 1888 GGCACAAATGAGCAAGCGCTTCCGAATTTGATAAATCCGAGAGGCGGAGCGGAC 1947
 QY 1441 AGTCGCTTCAACGATTTGAGCGTTTAAAGCGAG 1473
 DB 1948 AGTCGCTTCAACGATTTGAGCGTTTAAAGCGAG 1980

RESULT 7
 AR202319 2123 bp DNA linear PAT 20-APR-2002
 LOCUS AR202319
 DEFINITION Sequence 1 from patent US 6361986.
 ACCESSION AR202319
 VERSION AR202319.1 GI:20256858
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2123)
 Tiliy, Y., Eikmanns, B., Eggeling, L., Sahm, H. and Mockel, B.
 TITLE Process for the preparation of L-amino acids by fermentation and
 nucleotide sequences coding for the acdA gene
 JOURNAL Patent: US 6361986-A 1 26-MAR-2002;
 FEATURES
 source 1. 2123
 location/Qualifiers
 BASE COUNT 460 a 574 c 592 g 497 t
 ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 6.4e-293;
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGAGAACGCTTTTCCGACTATGTGTGGGCGCATGAAACACCTTCAGCATTTGACGCTC 60
 DB GTGGAGAACGCTTTTCCGACTATGTGTGGGCGCATGAAACACCTTCAGCATTTGACGCTC 567
 61 ATAGACTCGGTTTGGAGCCCTGACGCTTCACTTTCTTGGAACTGAACTCCCAATATGAC 120
 DB ATAGACTCGGTTTGGAGCCCTGACGCTTCACTTTCTTGGAACTGAACTCCCAATATGAC 627
 568 ATAGACTCGGTTTGGAGCCCTGACGCTTCACTTTCTTGGAACTGAACTCCCAATATGAC 627
 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGGCCAATGCGAT 180

628 AACCTCAATCAAGCTATGACAGACCTTGGAGCGGGCTGAAGCAAGGCCAATGCGAT 687
QY 181 GAATGGTAAATTAATGAGAGGACCCGTGAGGGGATTCGGGTAGCCGTTATTTGTC 240
Db 688 GAATGGTAAATTAATGAGAGGACCCGTGAGGGGATTCGGGTAGCCGTTATTTGTC 747
QY 241 GATTTTCTCTCTCGGCGGTTCTTTGGGACAGTCCGTCGGTGCATCATGAAGGG 300
Db 748 GATTTTCTCTCTCGGCGGTTCTTTGGGACAGTCCGTCGGTGCATCATGAAGGG 807
QY 301 ATTCAACCGGCGACAGAGCTGAAACTCCAGCTGTGTCCTCCGTCCTCCGGTGTG 360
Db 808 ATTCAACCGGCGACAGAGCTGAAACTCCAGCTGTGTCCTCCGTCCTCCGGTGTG 867
QY 361 CGCATGCGAGGAGAAATGAGCTTTTGTGATGATGATGATGATGATGATGATGATG 420
Db 868 CGCATGCGAGGAGAAATGAGCTTTTGTGATGATGATGATGATGATGATGATGATG 927
QY 421 CSTCAACCGGCGAGGAGCTGTTGCGGTCTGATGATGATGATGATGATGATGATG 480
Db 928 CSTCAACCGGCGAGGAGCTGTTGCGGTCTGATGATGATGATGATGATGATGATG 987
QY 481 GCCATGCGCTCGTGGGTTCACTGAGGATCTGATGATGATGATGATGATGATGATG 540
Db 988 GCCATGCGCTCGTGGGTTCACTGAGGATCTGATGATGATGATGATGATGATGATG 1047
QY 541 GGTTCCTGCGGTCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 1048 GGTTCCTGCGGTCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 1107
QY 601 CAGCAGCGGAGAAATTTGGTGAATACTGATGATGATGATGATGATGATGATGATG 660
Db 1108 CAGCAGCGGAGAAATTTGGTGAATACTGATGATGATGATGATGATGATGATGATG 1167
QY 661 TTGCGTGAAGCGGTGGCAAAACCTCAAAGTTATTCAGCCGTTAGAGGCAACGATG 720
Db 1168 TTGCGTGAAGCGGTGGCAAAACCTCAAAGTTATTCAGCCGTTAGAGGCAACGATG 1227
QY 721 TTTTCCCAACAACCTCTGGGCTGGGCACTTCGGGATGAGGAGGATGCGGCTCTG 780
Db 1228 TTTTCCCAACAACCTCTGGGCTGGGCACTTCGGGATGAGGAGGATGCGGCTCTG 1287
QY 781 GACCCGAGAGGCTGGAATCGGGAGATTAATGAAAGTTGGGGAGAGAGCTGCTCAAG 840
Db 1288 GACCCGAGAGGCTGGAATCGGGAGATTAATGAAAGTTGGGGAGAGAGCTGCTCAAG 1347
QY 841 CTTTCTGCTGCGGTGCTGGCGCATTTAGACCCGCTGTGCGCTGTCCTGCGCGCATC 900
Db 1348 CTTTCTGCTGCGGTGCTGGCGCATTTAGACCCGCTGTGCGCTGTCCTGCGCGCATC 1407
QY 901 GGGGGCGGGCCGCTGCTGATTTGGGAGATGCGCGCTTCAAGCTTGGGCGGAGAG 960
Db 1408 GGGGGCGGGCCGCTGCTGATTTGGGAGATGCGCGCTTCAAGCTTGGGCGGAGAG 1467
QY 961 CTGCGTTTTGCGCTGCTGCGCATTTGCTGCGCGCGAGCTTAACTCCGATCGTCTC 1020
Db 1468 CTGCGTTTTGCGCTGCTGCGCATTTGCTGCGCGCGAGCTTAACTCCGATCGTCTC 1527
QY 1021 ATCATGACACTCCGGCGCCGAATTTGTCGAGGCGGCTGAGGAGCTCGGCAACG 1080
Db 1528 ATCATGACACTCCGGCGCCGAATTTGTCGAGGCGGCTGAGGAGCTCGGCAACG 1587
QY 1081 TCGATTGGCGGACCTTGCCAAAGCTTATCCAGCGCTCCGCCCAACCGTTCCGTCATT 1140
Db 1588 TCGATTGGCGGACCTTGCCAAAGCTTATCCAGCGCTCCGCCCAACCGTTCCGTCATT 1647
QY 1141 ATTGCTCAAGGCGTGGCGGTGGCGCGCTGAGCATGCTGCGCGCATGCTTACCGG 1200
Db 1648 ATTGCTCAAGGCGTGGCGGTGGCGCGCTGAGCATGCTGCGCGCATGCTTACCGG 1707
QY 1201 GCCGAAAACGGTGGCTGTCCGATTTGCCACCAAGGGCGCTCGGCAATCTCTTCCG 1260
Db 1708 GCCGAAAACGGTGGCTGTCCGATTTGCCACCAAGGGCGCTCGGCAATCTCTTCCG 1767

QY 1261 GACACCAACGAGCGCGGGAATTCATAGAGGACAGGCGGTGAGCGGACGCACTTTTA 1320
Db 1768 GACACCAACGAGCGCGGGAATTCATAGAGGACAGGCGGTGAGCGGACGCACTTTTA 1827
QY 1321 AGCCAAAGGCTTATCGAGCGGATGCTGCGCAACCGAGACCTTTGTTGAAGAAATTC 1380
Db 1828 AGCCAAAGGCTTATCGAGCGGATGCTGCGCAACCGAGACCTTTGTTGAAGAAATTC 1887
QY 1381 GGCAATCAAGCAAGCGCTCTCCGAATTGGATTAACATCCGAGAGGGCGGACGCGAC 1440
Db 1888 GGCAATCAAGCAAGCGCTCTCCGAATTGGATTAACATCCGAGAGGGCGGACGCGAC 1947
QY 1441 AGTGGTTCAACAGATTGAGCGGTTAGCGGAG 1473
Db 1948 AGTGGTTCAACAGATTGAGCGGTTAGCGGAG 1980

RESULT 8
BD004741
LOCUS
DEFINITION
BD004741
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BD004741 2123 bp DNA linear PAT 31-JAN-2002
Replicable DNA, amino acid sequence, Corynebacterium microorganism,
shuttle vector, and process for producing L-amino acid.
BD004741
BD004741.1 GI:18632702
JP 2001008693-A/1
Corynebacterium glutamicum
Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
1 (bases 1 to 2123)
Tilke, I., Eggering, L., Eickmans, B., Zamu, H. and Meck, V.
Replicable DNA, amino acid sequence, Corynebacterium microorganism,
shuttle vector, and process for producing L-amino acid
Patent: JP 2001008693-A 1 16-JAN-2001;
DEGUSHA HUELS AG, FORSCHUNGSZENTRUM JUELICH GMBH
OS Corynebacterium glutamicum
PN JP 2001008693-A/1
PN 16-JAN-2001
PF 24-MAY-2000 JP 2000153547
PR 27-MAY-1999 DE 19924365.4
PI IVENNE TILKE, IOTHAR EGGERING, BERNHARD EICKMANS, HERMANN ZAMU,
PI VETLYNA MECKEL
PC C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, C12P13/10, PC
C12P13/08
PC C12P13/12, C12P13/20, C12P21/02, (C12N15/09, C12R1:15), (C12N1/21,
PC C12R1:15),
PC C12P13/04, C12R1:15), (C12P13/06, C12R1:15), (C12P13/08, C12R1:15), PC
(C12P13/12, C12R1:15), (C12P13/20, C12R1:15), (C12P21/02, C12R1:15), PC
C12N15/00
PC (C12N15/00, C12R1:15)
CC
CC
FH Key Location/Qualifiers
FT gene (508). (1980).
FEATURES
source
1. 2123
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"

BASE COUNT 460 a 574 c 592 g 497 t
ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 2123;
Best Local Similarity 100.0%; Pred. No. 6, 4e-293;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAGGAGGTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
Db 508 GTGAGAGGAGGTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 567
QY 61 ATAGACTCGGTTTGAACCTCGACAGCTTCAATTTGGAATGAACTCCCAATATGAC 120

Db 568 ATBAGTCGGTTTGGACCCCTGACAGCTTCATTCTTGGAAATGAACTCCCAATATGAC 627
 Qy 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTGAAAGCAAGGCCAAATGCAAT 180
 Db 628 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTGAAAGCAAGGCCAAATGCAAT 687
 Qy 181 GAATCGGTAAATTAATGAGAGAGCAACCTGAGAGAGGCAATCCGGTAGCCGTTATTTTGTCC 240
 Db 688 GAATCGGTAAATTAATGAGAGAGCAACCTGAGAGAGGCAATCCGGTAGCCGTTATTTTGTCC 747
 Qy 241 GATTTTCTCTCTCGGCGGCTTCTTGGGCAAGGTGCGTGGTGCGCATCATGAAGGCG 300
 Db 748 GATTTTCTCTCTCGGCGGCTTCTTGGGCAAGGTGCGTGGTGCGCATCATGAAGGCG 807
 Qy 301 ATTCACCGGCGCACAGAGCTGAAATCCCACTGCTGCTCCCTGCTCCGTTGGTGGCG 360
 Db 808 ATTCACCGGCGCACAGAGCTGAAATCCCACTGCTGCTCCCTGCTCCGTTGGTGGCG 867
 Qy 361 CGCATGCGAGGAAGACATGAGCTTTTGTATGATAGTGTGCTCAATACCGCGCTGTGCG 420
 Db 868 CGCATGCGAGGAAGACATGAGCTTTTGTATGATAGTGTGCTCAATACCGCGCTGTGCG 927
 Qy 421 CGTCAACCGGAGCGCGCATTTGCCCTTCTGCTGATTTTGGCGAATCCACGATGGGTGCG 480
 Db 928 CGTCAACCGGAGCGCGCATTTGCCCTTCTGCTGATTTTGGCGAATCCACGATGGGTGCG 987
 Qy 481 GCCATGCGCTCGTGGGGGTTTCATCTGGGGCATCTCATTTTGGGGAACCGCGCGCGAGATA 540
 Db 988 GCCATGCGCTCGTGGGGGTTTCATCTGGGGCATCTCATTTTGGGGAACCGCGCGCGAGATA 1047
 Qy 541 GGTTCCTGAGGTCCTCGCGTGTGTGAGTTAACCACTGAGGAGATGCGCTTCAGACGGTGTG 600
 Db 1048 GGTTCCTGAGGTCCTCGCGTGTGTGAGTTAACCACTGAGGAGATGCGCTTCAGACGGTGTG 1107
 Qy 601 CAGCAGCGGAGAAATTTGTGTAAACTGTGTGATGATGAAATTTGTGTGCGCACTCCA 660
 Db 1108 CAGCAGCGGAGAAATTTGTGTAAACTGTGTGATGATGAAATTTGTGTGCGCACTCCA 1167
 Qy 661 TTGCGTGCAGGCGGTGGCAAAACCTCAAGGTTATTCAGCGGTGAGGCAACGAGTGGT 720
 Db 1168 TTGCGTGCAGGCGGTGGCAAAACCTCAAGGTTATTCAGCGGTGAGGCAACGAGTGGT 1227
 Qy 721 TTTTCTCCAAACAATCTGCGCGTGTGCACTTCCGATGATGAGGCGATTTGCGGTTCTCGT 780
 Db 1228 TTTTCTCCAAACAATCTGCGCGTGTGCACTTCCGATGATGAGGCGATTTGCGGTTCTCGT 1287
 Qy 781 GACCCGCGAGAGGCTGTGAATCGGGAGATTAATGAAACGTTGGGGGAGACGTCGTCAAG 840
 Db 1288 GACCCGCGAGAGGCTGTGAATCGGGAGATTAATGAAACGTTGGGGGAGACGTCGTCAAG 1347
 Qy 841 CTTTCTGATGCGCGGTGGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGGCAATC 900
 Db 1348 CTTTCTGATGCGCGGTGGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGGCAATC 1407
 Qy 901 GGGGGCGCGGCGCTGTGATTTGGGAGAGATTCGCGCTTACGCTTGGCGCGGAGAG 960
 Db 1408 GGGGGCGCGGCGCTGTGATTTGGGAGAGATTCGCGCTTACGCTTGGCGCGGAGAG 1467
 Qy 961 CTGCGTTTGGCGCGTGTGAGCATTTTCGTGGCGCGAGCTTAACCTGCGCATGCTGTCC 1020
 Db 1468 CTGCGTTTGGCGCGTGTGAGCATTTTCGTGGCGCGAGCTTAACCTGCGCATGCTGTCC 1527
 Qy 1021 ATTCATCGACACTTCGGGCGCGAATTTGTGCGAGGGGCGTGAAGAGCTGGGCAATCGCAAGC 1080
 Db 1528 ATTCATCGACACTTCGGGCGCGAATTTGTGCGAGGGGCGTGAAGAGCTGGGCAATCGCAAGC 1587
 Qy 1081 TCGATTGGCGGCACTTGTCAAGCTTATCGAGCGTCCCTCCCAACGCTTTCGGTCAAT 1140
 Db 1588 TCGATTGGCGGCACTTGTCAAGCTTATCGAGCGTCCCTCCCAACGCTTTCGGTCAAT 1647
 Qy 1141 ATTGGTCAGGGCGTTGGCGGTGGCGCGCTGAGCCATGCTGCGCGGATCTGGTCAAGCG 1200
 Db 1648 ATTGGTCAGGGCGTTGGCGGTGGCGCGCTGAGCCATGCTGCGCGGATCTGGTCAAGCG 1707

Qy 1201 GCCGAAACGCGTGGCTGTCCGATTTGCCACGAGAGGGGCGCTCGGCCATCTCTTCCG 1260
 Db 1708 GCCGAAACGCGTGGCTGTCCGATTTGCCACGAGAGGGGCGCTCGGCCATCTCTTCCG 1767
 Qy 1261 GACACCAACGAGCGCGGAAATCATAGAGGAGAGGCGTGCAGCGGAGCGCATTTTA 1320
 Db 1768 GACACCAACGAGCGCGGAAATCATAGAGGAGAGGCGTGCAGCGGAGCGCATTTTA 1827
 Qy 1321 AGCCAGGCGTTATGAGAGGAGATCGTCCGCGAACCAGGACATTTTGTAAAGAAATTC 1380
 Db 1828 AGCCAGGCGTTATGAGAGGAGATCGTCCGCGAACCAGGACATTTTGTAAAGAAATTC 1887
 Qy 1381 GGCACCAATCAGCAAGCGCTTCGCCAATTTGATTAACAATCCGAGAGGGGCGGAGCGGAG 1440
 Db 1888 GGCACCAATCAGCAAGCGCTTCGCCAATTTGATTAACAATCCGAGAGGGGCGGAGCGGAG 1947
 Qy 1441 AGTCGCTTCACAGATTGAGGCGTTAGCGGAG 1473
 Db 1948 AGTCGCTTCACAGATTGAGGCGTTAGCGGAG 1980

RESULT 9
 AP005276/3 332050 bp DNA linear BCT 08-AUG-2002
 LOCUS Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 3/10.
 DEFINITION AP005276 BA000036
 ACCESSION AP005276.1 GI:21323419
 VERSION AP005276.1 GI:21323419
 KEYWORDS
 SOURCE Corynebacterium glutamicum ATCC 13032
 ORGANISM Corynebacterium glutamicum ATCC 13032
 Bacteria; Actinobacteria; Actinobacteriales; Corynebacteriaceae; Corynebacterium.

REFERENCE
 1 Nakagawa, S.
 Complete genomic sequence of Corynebacterium glutamicum ATCC 13032
 JOURNAL Unpublished
 TITLE 2 (bases 1 to 332050)
 AUTHORS Nakagawa, S.
 DIRECT SUBMISSION
 JOURNAL Submitted (24-May-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories; 3-6-6, Asehi-machi, Machida, Tokyo 194-8533, Japan (E-mail: snakagawa@kaken.kyowa.co.jp, Tel:81-44-829-3011, Fax:81-44-813-1611)
 COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.

FEATURES
 source
 1..332050
 /organism="Corynebacterium glutamicum ATCC 13032"
 /mol_type="genomic DNA"
 /strain="ATCC 13032"
 /db_xref="taxon:196627"
 /note="ATCC 13032"
 314..2518
 /gene="Cgl0654"
 /name="PF02554:Carbon starvation protein CseA"
 /codon_start=1
 /transl_table=1
 /product="Carbon starvation protein, predicted membrane protein"
 /protein_id="BAB98047.1"
 /translation="MKKSPKMSPTARVGLVFGVIAVWGAIASFNRGETINSVVLVAAVSIIASFVRLIEKYKVPKDRATPAEYVNGKDVLPDRRLFGHHPAIA GAGPLNGPWAOMGYLPBETLITLGVIPAGAVDYLVLTWSTRRSGSLQMRDEM GTVCGAGITLITISIMITITLIVALLIVNLAADSPWGFSTITLPIILFNGVDRLRPGKRVETSIIGVALLLAIVAGVADTSWGEVFTWSKTTLLALIGVIMAILP VMLLAPRDYISITPMKIGVIGLAVGLIFARPEVOMSPVSPFALEGNPVSGLIFP LFIITACGALSGFALISGTPPLVKESSQRMIGYGMIMESFVMMALITVILD RHLVFSNAPLALNGDPATAAEMVNSIGLIGADITTEQJSEAESVESVETIIRITG

APTLAFGMSSEILSGFIGAGMKAFWYHFAIMEFALFILTVDAGTRVARFNMVDTLGN
VGLRRFPDPSWITGVNMTISTVEVCALMGAILTMGTDPDLGGINVLPLFEGIANOLIAA
IALALIVIVVVKGIYKWMIPAVPLAIDLITVTMTRASMOKIEHSDPAICWMAONANFR
DAKSQGLTEFGAKSPKPAIDAVIRNTMIGILSITIPAVLVVYGAALAVICIKSRAR
AAGTPLETTEEDTESSEFPAPTGFLASSRDXEVMWDERYPGAPVSSGGH"
2522. .2725
/gene="Cg10655"
/gene="Cg10655"
2522. .2725
/gene="Cg10655"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98048.1"
/db_xref="GI:21323421"
/translation="WMALTHALMKIPRAVWVYLTELMDGTAYSKVYVHLKHHNDARI
PTERYAKTADODANPARGC"
complement(3092. .3703)
/gene="Cg10656"
complement(3092. .3703)
/gene="Cg10656"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98049.1"
/db_xref="GI:21323422"
/translation="MSNVINILVLMQAHSTHOVNLIPSPDATDQIILATPSILCCSQ
NRNVISWGVPRSEPRMEIKLTIRSAVSKRSPRGITSMOTENRCPTPGKMTMPLC
HRTGRAMNSTOHSCLHRCIYDNDVDDPLFSRSLNMVGCASPLVSQVILLBELFR
RTVGAFVFNHGMVTPSORINVAHTNLSNVGHCKSCVIL"
3750. .4706
/gene="Cg10657"
3750. .4706
/gene="Cg10657"
/codon_start=1
/transl_table=11
/product="Unclassified protein involved in propionate
catabolism"
/protein_id="BAB98050.1"
/db_xref="GI:21323423"
/translation="WMNTITQALIGQALHTTTRTSRKGAISSMKAPAPARAKMSIE
AYDRMRGEASPIWESGEDVIAWLISGLDITVITPLPARGEARALIDVYTESHA
EVSQAPIDILARSMGKLAQGLDLDVDSIVLHTSHTHVIGTGSNDPKPFEDSA
RETLDSIMYIFVALIEDRAMHERSVAPERARRETEILEMKISTVEDPMTTRYHS
VDPAKARGARAVITFEKDTVEDELAAVADAPLGRPAREOYIQKFRTLAEGVASE
KRODFLDAORTHELDELSEINLELDADILAKAVIEGLEF"
4715. .5629
/gene="Cg10658"
4715. .5629
/gene="Cg10658"
/codon_start=1
/transl_table=11
/product="per phosphonomutase and related enzymes"
/protein_id="BAB98051.1"
/db_xref="GI:21323424"
/translation="MSSAVAPTERRKALRALAPEIARMGAFSPPLARABOEGAF
EGVYGAVVADLALPDIGLTTTEVARSROIARVDLPVLVADGFGEPMSAAR
TYSLEADGAVAGCHLEDVNPBKRCGLHCKEYVGTGDIVRRIAAVNERRBOQFICA
RTDLAAGVSGIDSAITRAKAYADAGDMITFTEALISPAUPFKRAAVDIPLANHTEG
KTELLPAQLELDIGNAVIYPTLLIRIMGQVEALGDIANTGIQTDVDRNQHRSRL
YELLRYENANADQOVFTYSADSKYKIP"
5658. .6809
/gene="Cg10659"
5658. .6809
/gene="Cg10659"
/EC_number="4.1.3.7"
/note="PF00285:Citrate synthase"
/codon_start=1
/transl_table=11
/product="Citrate synthase"
/protein_id="BAB98052.1"
/db_xref="GI:21323425"
/translation="MSSATTDVVRKGLYGVADYTAVSKVMPETNSLTYRGAYVEDLV
ENCSEFEVYLLMHGELPTAQOLAEFNERGRSRLDAGLLSLHSLPKEMHPDMVNR

TAVSYMGTKDSEYFTTDSERIHKVGTLLAQLPMVLAMIRRRKGLDIIADDSKSYA
ENLISWFGEPSPGASNPADVRDPKESLLIYVHSPNASTFPAVITTSKSYSAI
TGALIGKPLHGGANFVNMTHALIDDPKAAAMINNALDNKVVWFGHRYKRGD
SRVDSMKSFRLEIARBDGEKVAMENMDAMDARPGIDPNDLDFPGAPVYHLGFPV
DFPTPLVIARVAGMTHVIEQYENNSLIRPLSEYNGEORVAPIERK"
complement(6846. .7040)
/gene="Cg10660"
complement(6846. .7040)
/gene="Cg10660"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98053.1"
/db_xref="GI:21323426"
/translation="MVEFVSQDVDELHFSFFELYDQAALDAHRCAHPHYVREAVA
KHVLPQSQKXNVGTRMLHNG"
complement(7129. .8172)
/gene="Cg10661"
complement(7129. .8172)
/gene="Cg10661"
/EC_number="1.1.1.82"
/note="PF02615:Malate/L-lactate dehydrogenase"
/codon_start=1
/transl_table=11
/product="Malate/L-lactate dehydrogenases"
/protein_id="BAB98054.1"
/db_xref="GI:21323427"
/translation="MPEVTYNAOQILVLCITDILTKGVPAADAVLVGDSIVQADIMWH
PSHGVLLPMYVRRLHSGANTTAHVEVLNDLGAVALIDGHNGIGQYLADHARKREAVT
RAMVFGIYAVSVRNSNHGTAMYTRKAAQGCVSIIITVASPMAWGRERKIGTN
PWSIAPFGGTATVVDIANTAVARGLYHARQNMPIPETWALTSBGAPTTDPAEAIN
GVVLPMAGHKGVASIFMMDVLSGVLTSSQSTKHGYPDPTPGAGALHFLDALVAIF
RDPQDFDASDLVGEVKSPTPKAQNTBEIYIPGSESDRARKNASAGHSIPEKTMEL
QELAIENHVYTHR"
complement(8265. .8987)
/gene="Cg10662"
complement(8265. .8987)
/gene="Cg10662"
/note="PF00392:Bacterial regulatory proteins, gntR family"
/codon_start=1
/transl_table=11
/product="Transcriptional regulators"
/protein_id="BAB98055.1"
/db_xref="GI:21323428"
/translation="MVQKQASPARTKADQVYTOIRREIEDTGNFGORMSEVWLVE
HTGASRTPYDALRLADRLITIERQAMVSPISRHIKDLEFPRRIYEVALLERI
SVGASKSPRIEFGESTIADPFELENSADADPTADRRLTSKFDLVVAANTHNOFIG
RSIILSLKPTTTRLRIIASHDARLROSVQEHIEWCEVASGDLRSAGACRQHLIIVE
KSIITLALINDSTSGQIDIRS"
complement(9094. .9798)
/gene="Cg10663"
complement(9094. .9798)
/gene="Cg10663"
/codon_start=1
/transl_table=11
/product="Hypothetical membrane protein"
/protein_id="BAB98056.1"
/db_xref="GI:21323429"
/translation="MRSRVLPALLITTAVSIPASAAITLADTKELCIASNTDSSA
VVVFPPDDEIDSVREQRDELDAQDPGKKAIESIADDDNAFPAELQVLDLASEB
GLAMLLPDDPTLADPNAEBSFKTEYDEAKDIISGSSDPASVLSQLOOAATGTR
TAETRAEVFADRTDDYVESQTLKEDPQNCIDAIIDARPIPLQYILIGAMIAVAIVL
GIRAWTNRKROSKHSQ"
complement(9878. .12094)
/gene="Cg10664"
complement(9878. .12094)
/gene="Cg10664"
/EC_number="1.1.1.42"
/note="TIGR00178:monomer_idh: isocitrate dehydrogenase,
NADP-dependent"
/codon_start=1
/transl_table=11
/product="Monomeric isocitrate dehydrogenase (NADP+)"

/protein_id="BAB98057.1"
 /db_xref="GI:21323430"
 /translation="NAK1ITRTDEAPLILATVSLKPVNEFAPATAGIEVETEDISLIG
 RIAOPERLTDOKVGNALIELGELAKPEANITLCPHISAPVPLKALIKELQOG
 YVILPELPAATDEEDDILARIYAVAGSAPVPLRFGNSDRRLPIAVKPFKPPHRM

Query Match 100.0%; Score 1473; DB 1; Length 332050;
 Best Local Similarity 100.0%; Pred. No. 3.4e-293;
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 60
 DB 213164 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 213105
 QY 61 ATAGACTGGTATTTGAGACCTTGACAGCTTCATTTCTTGAGTAAGAACTCCCAATATGAC 120
 DB 213104 ATAGACTGGTATTTGAGACCTTGACAGCTTCATTTCTTGAGTAAGAACTCCCAATATGAC 213045
 QY 121 AACCTCAATCAAGGCTATGACAGACCTTGAGCGGCTCGAAGCAAGCCAAATGCGAT 180
 DB 213044 AACCTCAATCAAGGCTATGACAGACCTTGAGCGGCTCGAAGCAAGCCAAATGCGAT 212985
 QY 181 GAATGGTATTAATCTGAGAGAGCGACCTTGAGAGGCGATTCGGTAGCCGTTATTTGTCC 240
 DB 212984 GAATGGTATTAATCTGAGAGAGCGACCTTGAGAGGCGATTCGGTAGCCGTTATTTGTCC 212925
 QY 241 GATTTTCTCTCGGCGGCTCTTTGGGCGAGCTGCGTGGTGCGCATCATGAAGGCG 300
 DB 212924 GATTTTCTCTCGGCGGCTCTTTGGGCGAGCTGCGTGGTGCGCATCATGAAGGCG 212865
 QY 301 ATTCAACCGGCGACAGAGCTGAAGCTCCGACTGCTGCTCCCTGCTTCCGCTGGTGG 360
 DB 212864 ATTCAACCGGCGACAGAGCTGAAGCTCCGACTGCTGCTCCCTGCTTCCGCTGGTGG 212805
 QY 361 CGCATGCGAGAGAGACATGAGCTTTGTTCATGATGCTGTCCATTAACCGGCGCTGTGCG 420
 DB 212804 CGCATGCGAGAGAGACATGAGCTTTGTTCATGATGCTGTCCATTAACCGGCGCTGTGCG 212745
 QY 421 CGTACCGGAGAGCGCATTTGCGGCTCTCGGTGATTTGGCGAATCCCAATGGGTGGC 480
 DB 212744 CGTACCGGAGAGCGCATTTGCGGCTCTCGGTGATTTGGCGAATCCCAATGGGTGGC 212685
 QY 481 GCCATGAGCTGCTGGGCTTCACTGAGGACATCTCACTTTTGCAGAACCCGCGCGCAGATA 540
 DB 212684 GCCATGAGCTGCTGGGCTTCACTGAGGACATCTCACTTTTGCAGAACCCGCGCGCAGATA 212625
 QY 541 GGTTCCTGCGGCTCTCGGCTGTGATTAACAACCTGAGGATCGCTTCAGACGCTGTG 600
 DB 212624 GGTTCCTGCGGCTCTCGGCTGTGATTAACAACCTGAGGATCGCTTCAGACGCTGTG 212565
 QY 601 CAGCAGGCGAGAAATTTGCTGAACCTGCTGATTTGATGAAATTTGTCTCCGACTCCCA 660
 DB 212564 CAGCAGGCGAGAAATTTGCTGAACCTGCTGATTTGATGAAATTTGTCTCCGACTCCCA 212505
 QY 661 TTGCGTGAAGGCTGAGCAAAAACCTCAAGGTTATTCAGCCGCTGAGGCAACGATGCT 720
 DB 212504 TTGCGTGAAGGCTGAGCAAAAACCTCAAGGTTATTCAGCCGCTGAGGCAACGATGCT 212445
 QY 721 TTTTCTCCAACAATCTCTGCGGTGACCTTCGCTGATGAGCGATTCGCGTCTCGT 780
 DB 212444 TTTTCTCCAACAATCTCTGCGGTGACCTTCGCTGATGAGCGATTCGCGTCTCGT 212385
 QY 781 GACCCGAGAGGCGCTGGAATCGGGAGATTATGAAACGTTGGGGCGAGAGCTCTCAAG 840
 DB 212384 GACCCGAGAGGCGCTGGAATCGGGAGATTATGAAACGTTGGGGCGAGAGCTCTCAAG 212325
 QY 841 CTTTCTGCTGCGCTGCTGCGCATTTGAGCCCGCTGTGCGGCTGCTCCGCGCGCATC 900
 DB 212324 CTTTCTGCTGCGCTGCTGCGCATTTGAGCCCGCTGTGCGGCTGCTCCGCGCGCATC 212265
 QY 901 GGGGGCGCGCGCTGCTGCTGATTTGGGAGATCGCCGCTTCAAGCTTGGGCGCGAGAG 960
 DB 212264 GGGGGCGCGCGCTGCTGCTGATTTGGGAGATCGCCGCTTCAAGCTTGGGCGCGAGAG 212205

QY 961 CTGCGTTTGGCGGTGCGGCAATTTGCTGCGCGCGGAGCTAACTCGCGATGCTGTC 1020
 DB 212204 CTGCGTTTGGCGGTGCGGCAATTTGCTGCGCGCGGAGCTAACTCGCGATGCTGTC 212145
 QY 1021 ATCATGACACCTCCGCGCGCGCAATTTGCTGACAGCGGCTGAGAGATCGGCATCGAAGC 1080
 DB 212144 ATCATGACACCTCCGCGCGCGCAATTTGCTGACAGCGGCTGAGAGATCGGCATCGAAGC 212085
 QY 1081 TCGATTGCGCGACCTTGTCCAAAGCTTATGACGCTCCCTCCCAACCGTTTGGTTCAT 1140
 DB 212084 TCGATTGCGCGACCTTGTCCAAAGCTTATGACGCTCCCTCCCAACCGTTTGGTTCAT 212025
 QY 1141 ATTGTGACAGGCGGTTGGCGGTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1200
 DB 212024 ATTGTGACAGGCGGTTGGCGGTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 211965
 QY 1201 GCCGAAACGCGTGGCTGTCCGCAATTTGACAGAGAGGCGCTCGGCAATCTTCCTCCG 1260
 DB 211964 GCCGAAACGCGTGGCTGTCCGCAATTTGACAGAGAGGCGCTCGGCAATCTTCCTCCG 211905
 QY 1261 GACACCAACACGCGCGGGAATCATAGAGCAAGGCGTGCAGGCGCAGCATTTTA 1320
 DB 211904 GACACCAACACGCGCGGGAATCATAGAGCAAGGCGTGCAGGCGCAGCATTTTA 211845
 QY 1321 AGCCAAAGGCTTATGAGAGGATTCGTCGCCCAACCGAGCACTTTGTAAGAAATTC 1380
 DB 211844 AGCCAAAGGCTTATGAGAGGATTCGTCGCCCAACCGAGCACTTTGTAAGAAATTC 211785
 QY 1381 GGCACCAATCAGCAAGCCCTTCGCAATTTGATTAACAATCGGAGAGGCGGAGCGCGAC 1440
 DB 211784 GGCACCAATCAGCAAGCCCTTCGCAATTTGATTAACAATCGGAGAGGCGGAGCGCGAC 211725
 QY 1441 AGTCGCTTCAACGATTTGAGCGTTTACGCGAG 1473
 DB 211724 AGTCGCTTCAACGATTTGAGCGTTTACGCGAG 211692

RESULT 10
 AX127145/c 349980 bp DNA linear PAT 11-MAY-2001
 LOCUS
 DEFINITION Sequence 7061 from Patent EP1108790.
 AX127145 AX114121
 ACCESSION
 VERSION AX127145.1 GI:14041133
 KEYWORDS

SOURCE
 ORGANISM Corynebacterium glutamicum
 Corynebacterium glutamicum
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE
 AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
 TITLE Novel polynucleotides
 JOURNAL Patent: EP 1108790-A 7061 20-JUN-2001;
 KYOMA HAKKO KOGYO CO., LTD. (JP)

FEATURES
 source location/Qualifiers
 1..349980

/organism="Corynebacterium glutamicum"
 /mol_type="genomic DNA"
 /db_xref="taxon:1718"
 /note="Seq 1 to long (3,309,400) split in 11, seq 7061
 0.600.001 0.949.980 ."
 BASE COUNT 79274 a 90638 c 98727 g 81341 t
 ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 3.4e-293;
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 60
 DB 281114 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 281055

Db	Accession	Source	Organism	Reference	Journal	Title	Authors
Db	279974	ATGCTCAGGGGCTTGGCGGATGCGCGCTGCGCATTCGCGCATTCGCTTACCGG	279915				
Qy	1201	GCCGAAAACGGGTGGCTTCCGCATTGCGACCCAGAGGGCGCTCGGCATCTCTTCCGC	1260				
Db	279914	GCCGAAAACGGGTGGCTTCCGCATTGCGACCCAGAGGGCGCTCGGCATCTCTTCCGC	279855				
Qy	1261	GACACCAACACAGCCCGCGAAATCATTAAGAGCACAAGCGCTGACGGCGACGACCTTTTA	1320				
Db	279854	GACACCAACACAGCCCGCGAAATCATTAAGAGCACAAGCGCTGACGGCGACGACCTTTTA	279795				
Qy	1321	AGCCAGGGGCTTATGACGGGATGCTGCGCGCAACCGACATTGTTGAAGAAATTC	1380				
Db	279794	AGCCAGGGGCTTATGACGGGATGCTGCGCGCAACCGACATTGTTGAAGAAATTC	279735				
Qy	1381	GGCAACATACACACAGCCCTCTCCCAATTTGATTAACAATCCGAGAGGGCGGACCGCAG	1440				
Db	279734	GGCAACATACACACAGCCCTCTCCCAATTTGATTAACAATCCGAGAGGGCGGACCGCAG	279675				
Qy	1441	AGTCGCTTACACAGATTGAGCGTTAGCGCAG	1473				
Db	279674	AGTCGCTTACACAGATTGAGCGTTAGCGCAG	279642				
RESULT 11							
AP005217/c							
LOCUS	300750 bp	DNA	linear	BCT 20-DEC-2002			
DEFINITION	Corynebacterium efficiens YS-314 DNA, complete genome, section						
ACCESSION	AP005217						
VERSION	AP005217.1						
KEYWORDS	GI:23492722						
SOURCE	BA000035						
ORGANISM	Corynebacterium efficiens YS-314						
REFERENCE	Corynebacterium efficiens YS-314						
AUTHORS	1 Kowabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y., Ikeo, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y., Usuda, Y. and Sugimoto, S.						
TITLE	The entire genomic sequence of Corynebacterium efficiens YS-314						
JOURNAL	Published Only in Database (2002)						
REFERENCE	2 (bases 1 to 300750)						
AUTHORS	Kawabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and Director-General of Biotechnology Center.						
TITLE	Direct Submission						
JOURNAL	Submitted (17-May-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, Nishihara 2-49-10, Shibusawa-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)						
COMMENT	Kawabayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan						
	Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan						
	Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan						
	Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan						
	Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan						
	The other authors are at the National Institute of Technology and Evaluation, Shibusawa-ku, Tokyo, 151-0066 Japan.						
FEATURES							
source	location/Qualifiers						
	1..300750						
	/organism="Corynebacterium efficiens YS-314"						
	/mol_type="genomic DNA"						
	/strain="YS-314"						
	/db_xref="taxon:196164"						
	252..620						
	/note="CE0885"						
	/codon_start=1						
	/transl_table=11						

/product="putative cold shock protein"
 /protein_id="BAC17705.1"
 /db_xref="GI:23492733"
 /translation="MPVSTVKKMYDPERGFVNSNGEDCVKQVLPKYTEIHOGQ
 RIEFDPAGRKGPALRVKLVLETPEPRRPHHTYPEELINGIISDMVTLEGTVPALLK
 GHYDPRKGEQVAKILRAVAKELBA"
 complement(13881..14408)

Query Match 40.9%; Score 602.4; DB 1; Length 300750;
 Best Local Similarity 63.9%; Pred. No. 5,8e-114;
 Matches 928; Conservative 0; Mismatches 521; Indels 3; Gaps 1;

21 TATGTGTGGGCGATGAAACACCTTCAGCTAGCTCAGCTAGAGCTGTTTGAACC 80
 26058 TAGGTGTAGAACCAAGACCAACCTCGCCAGGATCGATCGATGATGATGACAA 25999
 81 TGACAGCTTCATTTCTTGGAATGAACTCCCAATATGACAACTCAATCAAGCTATGC 140
 25998 GGACAGCTTCTGCTCTGGGACACTCCCGGAGTACGGGGAGATGATGACTACCG 25939
 141 AGAGACCTTGAAGCGGGCTCCGAGAGAGCCAAATCCGATGATGATGATGATGAGA 200
 25938 GAGAGCGCTTGGCCCGCCCGGAGAAATCCGGGGTGAATGAGGATGATGACGGGTGA 25879
 201 AGGACCGTGAAGGCGCATTCGGTAGCCGTTATTTTCCGATTTTCTCTCTCGGCG 260
 25878 GGGAAACGATGAGGGCTCCCGGTGGCGCTTCATCGCTCCAGTTCCGCTTCTCGGTG 25819
 261 TTTCTTGGGACAGTGGCGGTGGTGGCATCATGAAAGGCGATTCAACCGCGCACAGAGCT 320
 25818 TTGATTCGGTGGCCGACCTCCCGTGTCTCATCCAGGCGCATCCACCGGGCACCGGGA 25759
 331 GAAATCCACAGCTGCTCTCCCGTCTCCGGTGTGGCGGACAGGAGAAAGAAATCG 380
 25758 ACGTCTCCCGCTGTGATCTCCCGCTGCTGGGTGACACCGGATGACAGAGGACACCC 25699
 381 AGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 440
 25698 GCGCTTCCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25639
 441 GCGCTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
 25638 GCGCTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25579
 501 ATCTGGGATCTCACTTTTGGGAAACCGGCGCGAGATAGTTTCTCTGGGCTCGGCT 560
 25578 GCGCGGCACTTCACTTATGCGGACCGGATGATGATGATGATGATGATGATGATGATGAT 25519
 561 GGTGAGTTTAAACCTGAGCATGCGCTTCAGACGGTGTGACAGAGCGGAGAAATTTGGT 620
 25518 GGTGAGTGTGACCAACCGGACCCCATGCGGAGGGCGGTGACGCGGGGAGAACTCCG 25459
 621 GAAAACGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
 25458 GCGGACGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25399
 681 AACCTCAAGTTATTTACCGCGGTAGAGGACGATCGTTTCTTCCAAACATCTCTCG 740
 25398 GATTCATGAGGGTGTGGGGGAGGCGGTGCGCAGGACGCGCCGTGAGCTGCGGACCGG 25339
 741 CGTGCACCTTCCGCGTATGAGAGCGATTGCGGTTCTCGTGAACCGGAGAGCTGGAAT 800
 25338 GAAACACGCTGCTGCGTGGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 25279
 801 CGGAGAGATTATGAAACGTTGGGGGACAGCTGCTCAAGCTTTCTGTGCGCGTCTGG 860
 25278 GCGCGATCTGTGCGCGTCTGGGAGAAACGTTGACCTGTGCGGACAGGTGACGG 25219
 861 CGCATTTGAGCCCGGCTGCGGCTTCCCTGCGCGCATGCGGAGCGCGCGCTGCTGCT 920
 25218 TCGCATCTTCGCGGAGCAACGATGCGCTGCGCGCATGCGGAGCGCGCTGCTGCTGCT 25159
 921 GATTGGGAGAGA---TCGCGGCTTACGCTTGGGCGGAGAGGAGTGGCTTTGGCGGCTG 977

25158 CATGGGAGAGACCGTACCACTCCCGCTCGGGCGGCGGACATGGGTTTCGCCCGCG 25099
 978 TGGCATTTCCGTGGCGGCGGACGTAACCTTCCGATGTTGATCATTCACACCTCCGG 1037
 25098 CGGATGCGCCCTGGCCAGGGAGCTGACGCTCCGATATAGCATATGACACACCCCG 25039
 1038 CGCCGAATTGCGAGGCGGCTGAGAGCTCGGATCGCAAGCTCATATTGCGGACCTT 1097
 25038 CGCGAGCTGCGGGAGGCGGAGAGAAAGGCTGCGGTTTCATTCGCGGACTTT 24979
 1098 GTCCAGCTTATGACGCTCCCTCCCGACCGTTTGGTCAATTATGATGAGGCGTTGG 1157
 24978 GGGGAGCTTGTGACAGCTCCGAGTCCGACCGCTCCGATCATCTCGGACAGGCTGTGG 24919
 1158 CGGTGGCGGCTGGCCATGCTGCGCGCGGATCTGTGTAAGCGGCGGAAACGCTGGCT 1217
 24918 AGGGGTGCGCTGGCATGCTGCGCGGACCGGCTGCGGAGGAGAAACGCTGAGCT 24859
 1218 GTCCGATTGCGACAGAGGCGGCTCGGCACTCTTCGCGGACCAACACAGCGCG 1277
 24858 GTCCCGCTGCGCGGAGGCGGCTCGGCACTATCATCTTACCGTACACCTCCAGCGCC 24799
 1278 GGAATCATAGAGGACAGAGCGGTGACAGCGCACGACTTTAAGCAAGGCTTATGCA 1337
 24798 GAGATGATGAGACGAGAGGCGGTGTCGCGCATGATGATGATGATGATGATGATGATGAT 24739
 1338 CGGATGCTGCGCGAAGACGACCTTTGTAAGAAATTTCTGGGACATGACAGACG 1397
 24738 CGGACTATCGCGGAAACGATGATTTGATGAGGAGGTTGTGCGCAATTGACACAC 24679
 1398 CCTCTCGAATTGATTAACAATCCGAGAGGCGGAGCGGACAGTGGCTTACACAGATT 1457
 24678 CTTCCAGAAATTGAAACCAATCCGAGAGGCTGATCGGAAACAGGCTTACAGCACTA 24619
 1458 TGAAGCTTTAGC 1469
 24618 TGAAGCTTGC 24607

RESULT 12
 CGLT/G C 3013 bp DNA linear BCT 17-FEB-1995
 LOCUS C-glutamylc glut gene for citrate synthase and ORF.
 DEFINITION X66112
 ACCESSION X66112
 VERSION X66112.1 GI:505580
 KEYWORDS citrate synthase; glt gene.
 SOURCE Corynebacterium glutamicum
 ORGANISM Corynebacterium glutamicum
 Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 REFERENCE 1 (bases 1 to 3013)
 Authors Elkmann, B.J., Thum-Schmitz, N., Eggeling, L., Luedtke, K.U. and
 Sahm, H.
 TITLE Nucleotide sequence, expression and transcriptional analysis of the
 Corynebacterium glutamicum gltA gene encoding citrate synthase
 Microbiol. 140, 1817-1828 (1994)
 JOURNAL JOURNAL
 TITLE Direct Submission
 AUTHORS Elkmann, B.
 REFERENCE Submitted (15-MAY-1992) B. Elkmann, Institut f Biotechnologie 1,
 Forschungszentrum Juelich GmbH, I B T, Postfach 1913, 5170 Juelich,
 FRG

FEATURES
 source
 1. 3013 Location/Qualifiers
 /organism="Corynebacterium glutamicum"
 /mol_type="genomic DNA"
 /strain="ATCC13032"
 /db_xref="taxon:1718"
 prim_transcript 717..>2151
 gene 838..2151
 /gene="glt"
 838..2151
 CDS

```

/gene="glt"
/SC number="4.1.3.7"
/codon start=1
/translate table=11
/product="citrinate (6t)-synthase"
/protein_id="CAA46902.1"
/db_xref="GI:505581"
/db_xref="SWISS-PROT:P42457"
/translation="MFERDIAVATNNKAVLHYRGGEFEMDITIASRNNGVILGKLS
ETGLITPDGVSSTGTSKITYIDGDAILIRRGDIDIAENATFNNVSYILNGE
LTPPELHKFNEIRHHTLLDEDFKSQVNFPRDAHPMATLSSVITLTYQDQNP
LDEADQDKATVRLMAKVPMLAAVAHRKAGAPYPMDSLNARENFLRMVFGYPEPY
EIDPIVKALDKILLIADHQNOSTSTVMIGSAQANNFVSIAGINLSCGLHGA
NOAVLEMLDIEDISNMGDAEFENKKNEDGVRLMGFGHRYKAVKDPAAIYKETAH
EILEHGGDLDLDIAIKLEIALADYFISRLKYPVDVYTGILYVAMFPDPTFVL
PAIGRLPGWIAHRYBQLGAAGKINRPROVYGNESRKLVPRER"
2276..2632
/note="ORF"
putative binding protein"
/codon start=1
/translate table=11
/protein_id="CAA46903.1"
/db_xref="GI:50548"
/db_xref="SWISS-PROT:P42458"
/translation="MEKPOIELPVGPAPEPDLVSDIIVGEGAAEPGGEVHYVGV
FETGEPDSMDRGTSQFPLNGLIAGWQEGIPGMKVGRRQLTIPEAAVYGEBSGH
PLSGRTLVIIIDLSA"
BASE COUNT      716 a      818 c      799 g      680 t
ORIGIN
Query Match      25.3%; Score 372; DB 1; Length 3013;
Best Local Similarity 100.0%; Pred. No. 2.7e-66;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1102 AAGCTTATGACGCTCCCTCCCAACCGTTTGCGTATTGTCAGGCGCTTGCGG 1161
DB      3013 AAGCTTATGACGCTCCCTCCCAACCGTTTGCGTATTGTCAGGCGCTTGCGG 2954
QY      1162 GCGGCGCTGCGCATGTCGCCCGCATGTCGTCTAGCGCGCGAAGACCGTGTCTTC 1221
DB      2953 GCGGCGCTGCGCATGTCGCCCGCATGTCGTCTAGCGCGCGAAGACCGTGTCTTC 2894
QY      1222 GCATTGCCACAGAGGCGCCTCGGCATCTCTTCGCGACACCAACGCGCGCGAA 1281
DB      2893 GCATTGCCACAGAGGCGCCTCGGCATCTCTTCGCGACACCAACGCGCGCGAA 2834
QY      1282 ATCATAGACGACAAAGCGGTGACGCGACGCACTTTAAGCCAAAGGCTTATCGACGG 1341
DB      2833 ATCATAGACGACAAAGCGGTGACGCGACGCACTTTAAGCCAAAGGCTTATCGACGG 2774
QY      1342 ATCGTCGCCGAAACCGAGACATTTGTTGAAGAAATCTCTGCGCAATCGCAACGCCCTC 1401
DB      2773 ATCGTCGCCGAAACCGAGACATTTGTTGAAGAAATCTCTGCGCAATCGCAACGCCCTC 2714
QY      1402 TCCGAATTGATTAACAATCCGAGAGAGGCGGACCGGACAGTCGCTTACAGATTGAG 1461
DB      2713 TCCGAATTGATTAACAATCCGAGAGAGGCGGACCGGACAGTCGCTTACAGATTGAG 2654
QY      1462 CGTTAGCGCAG 1473
DB      2653 CGTTAGCGCAG 2642

```

```

RESULT 13
AE006979/c      AE006979      17910 bp      DNA      linear      BCT 27-APR-2001
LOCUS           Mycobacterium tuberculosis CDC1551, section 65 of 280 of the
DEFINITION      complete genome.
ACCESSION       AE006979 AE000516
VERSION         AE006979.1 GI:13880475
KEYWORDS
SOURCE          Mycobacterium tuberculosis CDC1551
ORGANISM        Mycobacterium tuberculosis CDC1551

```

```

REFERENCE
AUTHORS
1 (bases 1 to 17910)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouli, H.,
Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 17910)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouli, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..17910
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
145..1326
/gene="MT0918"
145..1326
/feature="MT0918"
/feature="identified by Glimmer2; putative"
/codon_start=1
/translate table=11
/product="hypothetical protein"
/protein_id="AAK45164.1"
/db_xref="GI:13880476"
/translation="MPSRAIVQBSDSYFPCNGRRPIMPKIVSVQSTRRLTSFV
GRKALNDVRLSDRLVTLTGPMKSRSLAQIQAIEFTYGGWDCDLAVTD
RDQVLSIMNLGIPVQGLAIDTLVAVINDARVTLVLDHEHLIDCAALIDSLR
SCPRLITLSTEAELGAGELTRVPLSLINDAIELEPRARVRSPFAIYADAVT
VGEICRRLDGVPALTEIARVTDTPVEIILAGNDRIVAGAGMVRPQTCAT
VQSHALSGEPRLHLRLAVFAGFPDIDGQAQVANDPEEGYTLGRFAELDYKAF
VVENNRGRAGRLILYVQYALEKLSGSDAVLARYRLKQPNQVVRAGSGVR
Y"
1358..2920
/gene="MT0919"
1358..2920
/feature="MT0919"
/feature="identified by Glimmer2; putative"
/codon_start=1
/translate table=11
/product="conserved hypothetical protein"
/protein_id="AAK45165.1"
/db_xref="GI:13880477"
/translation="MTAPPIAVERNTRSKVQOQADAVVALGRKPGILLCVBERFRAMD
LPMAADALFLMAETPRPLHVALVLSQPDNGRVLRFVSAVAARQOVAPMWR
RPHRSITLIGOWSMRTETEDLDYHVRSLAPRAGTAEALVALVEIHAQMIDRSRPL
WOVDLIEGLPGRCRVVYVHVALADGVSVMRLORITVADPHOROMPTLMEVPOAS
VAKHTAPRGSSRPILIAKGVIGARGVGMVRVADDTWRAQACGSGITLTAAPTPPL
NEPIAGARSVAGCFPIERLQVVAHADATINDVVLACGALRAYILISGALPEAPL
IAVNFQSLADTAVIDVFGQGNKIGTLMCSLATHLASVERLSAIRASMDGKAIA
GRSRQALAMSAALGAAPLALAMALGRVAPLRPPVITISNVPGQALYMGARLADAL
YLSAPVDGALNITSCGTNEQITFGTLCRRVAPALSIITDQLAHELELLVGSSEAG
PSTRRLRIAGRR"
3093..4388
/gene="MT0920"
3093..4388
/feature="MT0920"
/feature="similar to PID:2440108; identified by sequence
similarity; putative"
/codon_start=1

```

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

```
/transl_table=11
/product="citrate synthase"
/protein_id="AAK4516.1"
/db_xref="GI:13880478"
/translation="MADTDPTATLRYPGGEIDLOIVHATEGADIGALPPLAKGHTT
PDVGPANPAKSSITYIDGAGILRYGYPIDQIAEKSTPEVCYLLIYBLPDTDO
LAQFGRIORHMLHEDLKRFPDGPANHPVLSVNALSAAYODALPMDNGQV
EASTRLAKLPTIAAYAKSGOPPLYPDNSLTVENPLRLTFGPAPADPEV
VRALDMLFLIHADHEONCSTVRLVSGSRRLFTSISGGINALMGPHAGNOAIVL
MLEGIRSDGDVSEFRKYKREAGYKLMGSGHRKRYKYIDPRARIYKQADKIYALG
GDDSLGIKLEBEALTDYDFIERKLLPNDFYGLIYRALGFPTMFVLFALGRL
PGMIAMHEMDEGSKIQRPROIYTYTERDYVITDAR"
/gene="MT0921"
/complement(4429. .6036)
/gene="MT0921"
/complement(4429. .6036)
/sequence="similar to PID:1001724 PID:1001724, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, putative"
/protein_id="AAK45167.1"
/db_xref="GI:13880479"
/translation="MSDHRDFDVVVGGGNGLVAAAYLARAGLRVRLRLAQTG
AAVSIQADGVVALSRYSYLSLPSRIADLAPVRLARPFSSTYAPATAGRS
LILGTPBRAHLAIGAIPAHGPAFRRCRITVTAIMPILLIEPIRTBOARDI
VEYGEHAAAQMAVDPDIGHAIGAVANDLRCVITATDILGTPAHMPEISMONI
CFLHLVGSGTVMHVPITGMSVTSALATPAARGAETVTGADVPALDPDSTYHS
DGSDEHLVRGRFVLGVTPAVLASLGEPAVALPAGQVNNVVRRLPRLRDSV
TPQAFAGTFHVNESQDLDAVYSAQASGRPLDPLPCBAQCHSLTDPBILSLRUDG
AQTIVFGIHTPHSVFDEGLAERLTAVALINSVLAEPIDQVLTMDAQSKCIET
TTTTLDRLOTGNTGNIFFHAGLSWPFADNDPLDTPARQWGVATDHERIMLCGSGARR
GGAVSGIGHNAAMVLAICLASRRKSP"
/gene="MT0921"
/complement(6062. .6325)
/gene="MT0921.1"
/complement(6062. .6325)
/gene="MT0921.1"
/complement(6062. .6325)
/sequence="similar to Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45168.1"
/db_xref="GI:13880480"
/translation="MGKGRKPDSETLHRIPLVAEKALRAQLRHGISESEQQQL
RRIEILDQCMLLQRRALROTGGDPRAVVRADQVGYGYS"
/gene="MT0922"
/gene="MT0922"
/gene="MT0922"
/gene="MT0922"
/sequence="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="Ompa family protein"
/protein_id="AAK45169.1"
/db_xref="GI:13880481"
/translation="MAKRAIGOTPATTDARTOKYRSGRPMWLGAVVPIILIA
ICYGAFRRQSTYGTPTPTITPTSTGASLSLSLSISGCTVTLIDFPBEAA
KALMTALNGLAPGVNVIDIHPVVRSLDFSABEFTKSVIPDGLKVEEDIV
TLTGTPSBEHDAVKRAATSTPMKLVNNIEVGOAPGPGPAGPCADLSAINAV
TGGPIAFGNDGASLIPADYEILNRVADLKACPDARVINGTYTDMTGSEGINIPLSAQ
RAKIVADYLVAGVAGDHIAITGLSVNPIASNPATPEGRAKNRREIIVN"
/gene="MT0923"
/gene="MT0923"
/gene="MT0923"
/gene="MT0923"
/sequence="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45170.1"
/db_xref="GI:13880482"
/translation="MDFVIQWGCYLLAFLGSAVAVWVVTLSIKRASRDEGAABASA
AETGAQ"
```

```
gene
7575. .8105
/gene="MT0924"
7575. .8105
/gene="MT0924"
/sequence="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45171.1"
/db_xref="GI:13880483"
/translation="WMEHVMWLAGLAFITGMVLTSLMRRPVHQLVYKSVGSSA
KSKPTARKPAVKSPTARESPATAKTVAATESAAEQIPVAGEPAEPTPAAGEPAKI
PVVYAPYAGSGSARAGAGSGSPQWMLVKGRSDTRLVYTPDPTVPAQVWFQDEBS
AARAFPTPMRSTRRT"
/gene="MT0925"
/complement(8122. .9531)
/gene="MT0925"
/complement(8122. .9531)
/sequence="similar to PID:2440113; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAK45172.1"
/db_xref="GI:13880484"
/translation="MALGCCIPSAESDSCVCSGACMNIISRIIFARTPSLRRVVA
TALCAALPVLIGTVWVGITIDRKEHLDRIDBAAGFAIPVPRGIDEIPEVSPNDQ
ALITVRGNVYKNSDITTLPRQDDYADYVYRGVRYRVEYRGPSTVAVGATYD
ATVATENNHLIERVLLICTFAIAAVFAMLAFAVAFVQLAQTSTIDGDAFV
EVHSAEHLIAEAMRMLORINMEQRTEALASADFAVASHELRITLTARTNL
EVALSTLDPDORKEVINLDRSIRLEATLSALERLAQELSTSDHVPVDITDLQ
RAADRAIYEDLDSVSPSTCIIVGLPAGLRILAVNALANAVKAGATVQLASVS
SRACVETAIIDNCSGVPEGRQVYEFHFSGSTASHSGSGGLALVAQQLHGGTAS
LENSEPLGARVLRPSPS"
/gene="MT0926"
/complement(9473. .10183)
/gene="MT0926"
/complement(9473. .10183)
/sequence="similar to GP:2440114; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-binding response regulator"
/protein_id="AAK45173.1"
/db_xref="GI:13880485"
/translation="MGGMIDGVTSPRVLVVDDSDVLSLERGLRSGFEVATAVDGA"

Query Match 21.2%; Score 312.6; DB 1; Length 17910;
Best Local Similarity 54.0%; Pred. No. 3.7e-54;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

gene
CDS
65 ACTGGCTTTGGACCCCTGACAGCTTCATTTTGGAAATGAAACCTCCCAATATGACAAAC 124
11770 ACGCGGTGCTAGACCGGGGATCTTTGTCACAGCTAGGAGACCGCTGCGGTGCGG 11711
125 TCAATCAAGGCTATGACAGAGCCTTGAGAGCGGGCTCGAAGCAAGCCAAATGCGATGAT 184
11710 TAGCCGACTCTATGACGGGAGAGCTGCGCGCGCTCGGGCGGCCAACCGCGGACGCAAT 11651
185 CGTAATTAATCTGAGAAAGCAGCGTGAAGGGCATTTCCGTAGCCGTATATTTTGTCCGATT 244
11650 CGGTGACAGACCGGAGGAGGAGCGGATTCGGGCGGCGGGGTGCCGTGGTGTGAGT 11591
245 TTTCTCTCTCGCGGCTCTTTGGGACAGCGTGGCGTGGTGGCATCATGAAAGCGATTC 304
11590 TCGACTTCTCGGGGCGCTCGATTGGGAGGCGGCGGCGGCAAGCATCAACCGCCCGCG 11531
305 ACCGGCCACAGAGCTGAAATCTCCACTGCTGCTGCTCCCTGCTTCGGTGGTGGCGCA 364
11530 AGCGGCGACCGCGAGCGGCTGCGCTACTGCGGTGTCACCAAGCTCGGAGGACCGCGCA 11471
365 TGCAGGAAGACATCGAGCTTTTGTCAATGATGTGTCCATTAACCGCGGCTGTGAGCGTC 424
```



```

/db_xref="SWISS-PROT:Q10536"
/translation="MRIGVCTTDPARQAAVEAAGARDELGAAPSILAVLLGSRAR
TDRADVLAQVOMIDPVALVGTICAOAVAGHEIEDPAVVVLAAGLAETPOLDR
VRTGSGALITGRFPRTARADLLHLLPPTTFSNLLIEHPTDLEPTAVGVSNGCR
RGGDTRLEFHDVLTSGVAVGLPKMGVPIVSGCRPGIPYIVTGADJLITELGG
RPLRLRLEIVGLSPDERALVSHLQIGIVDEHIAAPGQDFVIRGLGADPTGS
IEIDVAVGATMOQOVDRADAGADDRLRTERAARLFGRAAGALLFTCNCRGRME
GAHDHASTIEELGIIPLAGFFAAGEIPIAGRNALHGFATSMALFVDME"
complement(1314, .1802)
/gene="Rv0875c"
/complement(1314, .1802)
/note="Rv0875c"
/gene="Rv0875c" (MTCY31.03c), len: 162, unknown, possibly
exported protein, highly similar to MLCB57.11 Mycobacterium
leprae cosmid B57, 033056 HYPOTHETICAL 18.1 KD PROTEIN
(162aa). FASTA scores: opt: 789 z-score: 963.4 E(): 0;
71.4% identity in 161 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0875c"
/protein_id="CAA97382.1"
/db_xref="GI:1314012"
/db_xref="SWISS-PROT:Q10537"
/translation="MKRGVATLPVILVILSVAGAAMLVLRGHGPOPEISAYSHG
HLTRVGYLCNVVDLDCCQPOAGELPVSERPVQLSPEVISRAPRLLQVODP
ANTSTLPRDRLAVTIPTVDPQGRGLGIIVQLTLVVDHSGELRVPHAEWSVRL
IF"
complement(1799, .3445)
/gene="Rv0876c"
/complement(1799, .3445)
/note="Rv0876c" (MTCY31.04c), len: 548, unknown; possible
membrane protein, highly similar to MLCB57.12
Mycobacterium leprae cosmid B57 033057 HYPOTHETICAL 61.7 KD
PROTEIN (579aa). FASTA scores: opt: 2850 z-score: 2863.4
E(): 0; 81.0% identity in 568 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0876c"
/protein_id="CAA97383.1"
/db_xref="GI:1314013"
/db_xref="SWISS-PROT:Q10564"
/translation="MAPTGRTRNGSVVGHGPMANYPDDANYSRRPPMPGSAR
YLPLEGEDEPERSRVPRTTRRGRITVTRAAARSRBMGRMVLVLRATATAGAD
KSGLTALMPWANPAVDASAMVALANTLFPAAAGSEKSRVALYILITIPRATIA
LIGPADRLQHRRVAPALSGLTALAVLIMYDGTGSPSWLYPCALAMVFS
KSFVLSAATVPMPPTIDLVANSRLVFGLGTLAAGVIAAGVFCVTLFQLP
GALFVVAITTAGAGISMRIPRVEVTSGEVGLSYHRDRGLRRRPEEVKNGGT
LRPLGRNITITLGNCTIKVWVGLFLYPAPVAKAHANGVQLGMLGIGAAGV
NAGNFTSRLQIGRPVAVVCTVAVLATAAAGSLATATATLITGSSALAK
ASDASLQHDLPBESRAGSFGSESTLQIAWLGAVGVLYTTELWGTFTVSSALLI
GLAQTVISFRGSLIPGLGGRPFVMAEDSTTKRGAAVAQ"
complement(3457, .3465)
/note="gagstgac, potential rbs upstream of Rv0876c"
3572..3575
/note="gagc potential rbs upstream of Rv0877"
3583..4371
/gene="Rv0877"
3583..4371
/gene="Rv0877"
/note="Rv0877" (MTCY31.05), len: 262, unknown,
equivalent to MLCB57.13 Mycobacterium leprae cosmid B57
033058 HYPOTHETICAL 28.2 KD PROTEIN (269 aa; E(): 0.80.5%
identity in 257aa overlap.)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0877"
/protein_id="CAA97384.1"
/db_xref="GI:1314015"
/db_xref="SWISS-PROT:Q10539"
/translation="MTGPTESAVATVADMEGLAVALRGADQRAAVERSGPEAV
GDYLGVSVEDGAATRFIAHLPGVQGMQAVVAVSYSGADVTSEVLVAGPALL
APDWVFWGQVAPGDLSPGDLAPAKDDRLVPGVYTAGADQADVTAEIGLRWVW

```

```

SAMRAOASORWHDGYPGSGAMARSTRVCRDQCFPLPLAGSLGAMFGVCNLSAD
GHVVDROYGCGASHDTTAPAGSGPIRVEPYDDVDLIDIEKPAES"
complement(4380, .5711)
/gene="PPE"
/complement(4380, .5711)
/gene="PPE"
/note="Rv0878c" (MTCY31.06c), len: 443, unknown; similar
to hsc6 MYCT0 P42611 hypothetical 50.6 kd protein in hsp65
3' region. (517 aa). FASTA scores: opt: 1044, z-score:
954.9, E(): 0, (47.4% identity in 397 aa overlap). Similar
also to MTW014_3; MTC165_2, MTC198_24, MTC1307_23,
MTCY48_17, MTW004_5, MTW004_3 etc."
/codon_start=1
/transl_table=11
/product="PPE"
/protein_id="CAA97385.1"
/db_xref="GI:1314016"
/db_xref="SWISS-PROT:Q10540"
/translation="MNFVLPPEVNSARIYAGAPAPMLAAVAMDGLAELGMAAS
FSLLISGLTAGPGSGAPPAAMAAAPVLSMNTATARAEGAAARAAVAYA
ARAATAPALVAANRNOLLSLVSNLPAATAEASVEQLMADVAMGVHGG
ASTVASQLTWQQLSLVLPVVTAPAGVAPALAIPLAGVNTGVNPLGIGNIG
NNVSGNTGDNFNGIGNANLGNIGNANLGNAGAFENFNGNDNTNFGSN
AGFLNTGSGNBSGNLGFNAGADNTGNSGDTNFGNSGDLNTGSPVTVGVAN
SGFNTGTGSGFNSGNSGSPONLNGSSGFNADSTSGFONAGTALTTRASSTWA
DSPRAMEIRAPSRLOVWRTARATTARECSIVIIISVSTGAPPKKVGNSG"
complement(5719, .5721)
/note="yga (or aeg) potential rbs upstream of Rv0878c"
complement(5989, .6264)
/gene="Rv0879c"
/complement(5989, .6264)
/note="Rv0879c" (MTCY31.07c), len: 91, unknown, equivalent
to MLCB57.14 Mycobacterium leprae cosmid B57 033059 HYPOTHETICAL 9.8
KD PROTEIN (91 aa), E(): 1.2e-25, 76.9% identity in 91 aa
overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0879c"
/protein_id="CAA97386.1"
/db_xref="GI:1314017"
/db_xref="SWISS-PROT:Q10541"
/translation="MSVENSOIREPPLPVLEWVPIAVGALAMLVAAVAFVVG
LASRPVTVAGLATTGLGTTIFVWQLAARGARQAQAGLETYIDPK"
6442..6873
/gene="Rv0880"
6442..6873
/gene="Rv0880"
/note="Rv0880" (MTCY31.08), len: 143, unknown, equivalent
to MLCB57.15 Mycobacterium leprae cosmid B57 (143 aa)
033060 HYPOTHETICAL 15.8 KD PROTEIN. FASTA scores: opt:
818 z-score: 980.3 E(): 0; 89.5% identity in 143 aa
overlap."
/codon_start=1

```

Query Match 21.2%; Score 312.6; DB 1; Length 37630;
 Best Local Similarity 54.0%; Pred. No. 3.4e-54;
 Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```

QY 65 ACTCGTTTGGACCCCTGACGCTTCTTCTTGGAATGAAACCCCAATATGACAAC 124
DB 35657 ACGGCGTCTAGACCGGGGATCTTTCCTCAGCTGGATAGAGACCCCTGCGGTGCGG 35598
QY 125 TCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGAAT 184
DB 35597 TAGCCGACTCTATGCGCGGAGCTGCGCGCTCGGCGGCGACCGCGCGGACGAAT 35538
QY 185 CCGTAATTACTGGAAGGCAACCGTGAAGGCAATCCGTAACCGTATTTTGTCCGATT 244
DB 35537 CGGTGACAGACCGGTGAAGGCAACCGTATTCGGGCGCGGCGGTGTGTCGCTGATG 35478
QY 245 TTTCCTTCCTGCGGCTTCTTCTTGGGACCGGTGCGGTGCGGATCATGAAGCGAATTC 304

```


Db	354177	TCGACCTTCTGGGGGGCTCGATTGGGGGTGGGAGGCGCGAAACGGATACACCGCGCGGTGG	354118
QY	305	ACCGGCGCACAGAGCTGAAATCCCACTGCTGGTCTCCCTGCTTCGGGTGGCGCGCA	364
Db	35417	AGCGGGCCACCGCGGAGCGGCTGCGCGTACTGGCGGTACCAAGCTCGGGAGGCAACCCGCA	35358
QY	365	TGCAGAAAGACAAATCGAGCTTTTGTCAATGATGGGTGCATATACCGGGCGTGTGCAAGTTC	424
Db	35357	TGCAGAAAGGACAGGTGCGCTTTTGTGCAATGATGGAAGATCGCTGGCGGCATCACTGG	35298
QY	425	ACCGGAGGCGCATTTGTGCGTCTCTGGTGTATTTGCGCAATCCACAGATGGGTGGCGCA	484
Db	35297	ACAACACAGGCGCGCTGGCCCTACTGGTCTATTTGGCGCATCCGACCAAGGGGTGAGTTT	35238
QY	485	TGCGCTGTGGGGTTTCATCTGGGCATCTCACTTTGCGGAACCGGCGCGCAGATAGTT	544
Db	35237	TGCGGTGTGGGGCTCGCTGGGCGATCTCAACCGTCGCGAGCGGGGCGCTGTAGCGCT	35178
QY	545	TCCGTGGGTCCTCGGCGGTGGAGTTAACATGCGGCAATGGCGTTCCAGACGGTGTCAAC	604
Db	35177	TTCTGGGACCAAGGGTCTATGATGTTGCTATATGGCGACCCCTTCCATTCGGGTCAAA	35118
QY	605	AGCGGAGAAATTTGTGTAAACTGTGTGATTTGATGAAATTTGTGCGCACTCAATATGC	664
Db	35117	CCGCGAGAAATCTACGCGCGGATGGATTCATCGACGCGCTGTGGCACTGGACCGGCTAC	35058
QY	665	GTGCAAGCGGTGGCAAAACCCTCAAGGTATATCAGCGGTAGAGGCAACGATCGTTTTT	724
Db	35057	GACCGATGCTGGATCTGCGTGTGACGGTGTCAATCGACGCTCCGAAACCGCTTCGGGAC	34988
QY	725	CTCCAAACAATCTCTGGGTGGCA---CTTCCGGTATGAGAGGGGATTTGGCGCTTCGTG	781
Db	34987	CGCAGACGCCCGCGCCCGTACCCGATGGCCACGTGGGACCTGGTGTGGCAATCCGCGC	34938
QY	782	ACCAGGAGGCTCTGGAATCGGGGAGATTAATGAAACGTTGGGGGCGAGAGTGTCAACG	841
Db	34937	GGCGGAGACGGCGCGGGCGTCAAGGACGCTACTGCGACACGGCGCACCGACCGGATTTGT	34878
QY	842	TTTCTGTGCGCGTGTGCGGCAATTGAGCCCGGTGTGCGCGTTCGCTGGCGCGCATG	901
Db	34877	TGTCAAGAACCGATCAAGGCGAAGCG---GCGAACACGCTGTGGCGCTGGCCCGTTTG	34821
QY	902	GGGGCCGGCCCGTGTGTCTGATTGGGAG-----GATGCGCGTTTACGC	946
Db	34820	GCGGCAACCCACGCTGTGTCTCGGCGCAGAAAGGCGAGTGGCGCGGGGAAAGCACG	34761
QY	947	TTGGGCGCGAGAGCTCGTTTTCGCGGTGTGGGCATTTTGCCTGGGCGCGAGCTAAAC	1006
Db	34760	TGCGGCGCGCTGCGCTTACGGAAGCCCGAGCGGAGATGCGCTTCGCGCGCGAATTTGCC	34701
QY	1007	TGCCGATCTGTTCATCATCGACACTTCGCGCGCGCAATTTGTGCGAGCGGCTGAGAGAC	1066
Db	34700	TGCGGCTGTGTCTGTTCATTGACGCGCGCGAGACCGGCTTGTGTGCGCGCGCAACAG	34641
QY	1067	TGCGCATCGCAAGTTCATTGCGCGCACCTTTGTCAAGCTTATGACGCTCCCTTCCCA	1126
Db	34640	GCGGCGTGGCGCGCGCAATGCGGATTTGCTTGGCGGAGCTGTCAAGCTGATATCCCGCA	34581
QY	1127	CCGTTTGGGTCATTATTTGTCAAGGCGGTGTGGCGGTGGCGCGCGCATGCTGCCGCGG	1186
Db	34580	CCGTGTGATCTGTCTGTGGCGCAAGGACAGCGCGCGCGCGGCGCTGGGAGATTTGCCGCGG	34521
QY	1187	ATCTGTCTACGCGCGCGAAACCGCGTGTCTCGCATTTGCCACCAAGAGGCGCTTCGG	1246
Db	34520	ACCGGCTGTGGCGCGCACTCAAGGCTGTGGGCGCTTGGCTCCCGAAGGAGCAGCG	34461
QY	1247	CCATCTCTTTCGGGACACCAACACGCGCGCGGAAATCATAGGCAACAAGGTGTGACG	1306
Db	34460	CGATCTGTTCGAGACATGTCTCATTCGCCCGCAACTCGCTGCGCGCCAAAGATCTCGGT	34401
QY	1307	CGCAGCACTTTTAAGCCAAAGGGCTTATGACGGGAATCGTCCGCGAAACGAGCACTTTG	1366
Db	34400	CGCGGCACTTACTAAGTGGGAGTTGTGCACACCAATCGTGCCTGGAGTACCCCGAGCGG	34341

Oy	1367	TTGACGA	1373
Dd	34340	CAGACGA	34334
RESULT 15			
BX248337/c			
LOCUS			
DEFINITION	327650 bp	DNA	linear BCT 11-JUN-2003
ACCESSION	BX248337		
VERSION	4/14		
KEYWORDS	Mycobacterium bovis subsp. bovis AF2122/97		complete genome; segment
SOURCE	4/14		
ORGANISM	complete genome.		
	Mycobacterium bovis subsp. bovis AF2122/97		
	Bacteria; Actinobacteria; Actinobacteriales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium tuberculosis complex.		
REFERENCE	1		
AUTHORS	Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H., Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Morenne,C., Simon,S., Harris,B., Atkin,R., Doggett,J.J., Mayes,R., Keating,U., Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and Hewinson,G.		
TITLE	The complete genome sequence of Mycobacterium bovis		
JOURNAL	Online Publication		
REMARK	PMNS 10.1073/pnas.1130426100 (Microbiology)		
REFERENCE	2 (bases 1 to 327650)		
AUTHORS	Garnier,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-MAR-2003) Garnier T., Unite de Genetique Molculaire Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone, Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unite de Genetique Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France		
FEATURES			
source	Location/Qualifiers 1..337650 /organism="Mycobacterium bovis subsp. bovis AF2122/97" /mol_type="genomic DNA" /strain="AF2122/97"		
gene	/db_xref="taxon:233413" 698..2170 /locus_tag="Mb0912"		
CDS	698..2170 /locus_tag="Mb0912" /note="Mb0912", -, len: 490 aa. Equivalent to RV0888, len: 490 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 490 aa overlap). Probable exported protein. Equivalent to AAK45157.1 from Mycobacterium tuberculosis strain CDC1551 (507 aa) but shorter 17 aa. Contains possible N-terminal signal sequence."		
gene	/codon_start=1 /transl_table=1 /product="PROBABLE EXPORTED PROTEIN"		
CDS	/protein_id="CAD93773.1" /db_xref="GI:31617664"		
gene	/translation="MDYARKIGVGALVALVGLGAAVTTHAISAAPDPSSSTSP VDCPSLGSASLSIAIPGASVPGVRQVDPCSIDDLNLIDPLAVRNGLVP I ENTPANPOQVSVEPGGVGPVRPACDPRNMFPAIREBA PGCPGHGYTVNQR TASFTIADPGFVGTDTFSVNNSDDTSLHVLHGAGLYGFPHGDVAIVTAVGNPT DTISGFSQMTLVNIAGLPPLSSAILPRFYTKIEIKRLNAVNVANVQDFAYHQFL I KSKMSQMPPEPIITLMPIGVFPSSGLNTLSFFKORLDROVMCTSDNCITLKG F TYSOMLPGADVYNLTHTNGCGEPTTNAAQVNYIQNSAGAVIVTGDFNARY SDPOSLLFOAGVNGLTDAWVOGEHPPTNPFPFCMVNCEBGLDKIFRRSGQVTL QAYSYNENARPKEPNSGEPILSHSPRAVNGFHVADNVAVR"		
gene	complement(2205..3326) /gene="cItA"		

gene

```
CDS
/locus_tag="Mb0913c"
complement(1205..3326)
/feature="CITA"
/locus_tag="Mb0913c"
/EC_number="2.3.3.1"
/note="Mb0913c, cita, len: 373 aa. Equivalent to Rv0889c,
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 373 aa overlap). Probable cita
(alternate gene name: gltA), citrate synthase 2 (EC
4.1.3.7), highly similar to others e.g.
CAB95899.1|ALJ59988 putative citrate synthase from
Streptomyces coelicolor (387 aa); P39119|CISY_BACSU
citrate synthase II from Bacillus subtilis (366 aa), FASTA
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa
overlap), etc. Also similar to Rv0896|MTCY31.24 from
Mycobacterium tuberculosis (29.2% identity in 274 aa
overlap) and Rv1131. Contains P500480 Citrate synthase
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."
/codon_start=1
/transl_table=1
/product="PROBABLE CITRATE SYNTHASE II CITA"
/protein_id="CAD93774.1"
/db_xref="GI:31617665"
/translation="MTVVBNFVPGLDGVAFTEIAPEDKXGALRYRGVIEDLYS
ORVTEGVWALLVDGNFGSLPPAPFPPLPHSGDVRYDVAQLAMLPWIGVAPLD
IDDAFARQOOLARASVMAISYVAQSRGIVOPAPORITDECSVTAPRMTWQEPDP
RHRIDAVWYVSAEHGMNSTFRPARYASTGADVAAALSGAIGMSPLIGCAPRY
LPMIDVERAGDASVKGITDRGKLMFGHRYRADPPRARIARAELTGARIE
VAVAEQALSELRERRPDRAIETIVERMAVVIDFAVPANMMPAMFTCGRTGWCA
HLEQRKGLVPSAIVYGPGRPESVDGMEVLTTA"
complement(3413..6061)
/locus_tag="Mb0914c"
complement(3413..6061)
/locus_tag="Mb0914c"
/note="Mb0914c", len: 882 aa. Equivalent to Rv0890c,
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 882 aa overlap). Probable
transcriptional regulatory protein, luxR family, highly
similar (but shorter 238 aa in N-terminus) to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
(generally in part) to others e.g. T50568 probable
multi-domain regulatory protein from Streptomyces
coelicolor (1334 aa); P10957|NMRL_ECOLI nitrate/nitrite
response regulator protein from Escherichia coli (216 aa),
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99
aa overlap); etc. Also highly similar to others from
Mycobacterium tuberculosis e.g. MTCY02B10.22, MTW008.44,
MTV036.21, and MTCY31.24. Contains P500017 ATP/GTP-binding
site motif A (P-loop) _P800622 Bacterial regulatory
proteins, luxR family, signature, and probable helix-turn
helix motif from aa 836 to 857 (Score 1559, +4.50 SD).
BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
REGULATORS."
/codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(PROBABLY LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:31617666"
/translation="MRALLAONRLVTLCTGGVGVKTRLAIOIASSELDLQCFVDLA
PITSGVIAATARAAGVLPDQGRSTMSLSRPTGNRMMLVNDNCHELLDACAAVY
ELDACEPLTLTASREPITGAGETTRVPSMSTIDDAVELPADRASVVGQPTTAN
NAAAVGEICRLDGLPLAIEFAARVSRMSPLELADGDDCFRLLAGSVRAVQOOT
LRASIDWSHALLTEEOILFRRLAPFVGGLPAAVRAVAAASDDPPEVLDQTLVLD
KSLVADDCOGRTRYLETETRYALVEVKGSGEADVADYARADYTLAAALNTPAD
DHORLVARETEIDNLRAPAFMSRENHITGALQASLSIOIIFGQRALRGSLSPNS
ILDEORFRLRIVASTARALADKAMSTWLTSPVATDIIAIPQOOLAAARRVQDP
AAVRAITACGSSGYNAEAPFAETMDARADDKMTTCQILTYMGVETTCISGDP
NALRAAEERDLADITIGDRFVSRSCTMWSLQWMAWNLBALSLEITAEVANS
DVPTKVLGLYQAOVLVCGASAAHAIAGACIAATEAGVYOGIGVAAVYALAA
DVTAALEASDARPLIRAQPOVTHVLMACAGALHITVQGNPDAMELLAGLAGEVSH
RMVALTRARVATARGPELARDADHAALACGALHITVQGNPDAMELLAGLAGEVSH
```

```

SEGVRLGAAALROOTROYRFXIKWDAGYOASVTALBEAMGDEDFDEAAGAALSTD
EAIYAOGRGERRRPARMGSLPTRESDVRLVSELSKDIARLFFVSPRTVQTHL
THYAKTGLASRVOLVDEARSGSPS"
complement(6063..6920)
/locus_tag="Mb0915c"
complement(6063..6920)
/locus_tag="Mb0915c"
/note="Mb0915c", len: 285 aa. Equivalent to Rv0891c,
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q1102|MTCY02B10.22
PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in
247 aa overlap); MTW036.21, MTW008.44; MTCY02B10.23. Also
shows similarity with several adenylate cyclases and
hydrolases from other organisms."
/codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:31617667"
/translation="MLEPNAVHNSLPNIDIDHALIRGEDHPPTCAKCAVGARISALGS
LDRBYHSLRYAAPDPVGRCEPVPRRRVTIANQGDVSLPTPGYTLTLADVEEST
HLNMCPEDEATNIAHLDHTVSEALITHGVQYKRYREGSPAAEFTRASDAACALD
LQRTSLAPILRLRGLTGEVOLDELYVGTITRTAALRLAAGGVLSAAATGDLV
GRLPADMLVLDLGRHPLRGLRPBREWQOLCHPDIREKFPRLRTAKSSPTSILPAQFTT
FVGRRAQIS"
7318..8805
/locus_tag="Mb0916"
7318..8805
/locus_tag="Mb0916"
/note="Mb0916", len: 495 aa. Equivalent to Rv0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monoxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monoxygenase from Pseudomonas aeruginosa (491 aa);
CAB95668.1|ALJ32674 monoxygenase from Streptomyces
coelicolor (519 aa); P12015|CWO_ACS cyclohexanone
monoxygenase from Acinetobacter sp. (542 aa), FASTA
scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to Rv0565c, Rv354c,
Rv3083, etc from Mycobacterium tuberculosis. Has
hydrophobic stretch at N-terminus."
/codon_start=1
/transl_table=1
/product="PROBABLE MONOOXYGENASE"
/protein_id="CAD93777.1"
/db_xref="GI:31617668"
/translation="MTGRCPTVAVVGAGSGMCVAITLLSAGITDVCTIYEKADVGT
WRDVTYGLTCDVPSRLYOYSPAKNPMTWOFMRGGEIOOYLGIARYGRLRIRRG
ATVYSAPFDGRVLTPTDSCSTSVLFLSATGLHPRILPPLAGLDDEPGTYFHSAR
WDHTVPLGRRIAVITGSGTVGVGLVGLAIVAGKVTMFGTAVOVLPMPPRRSKLAR
VFHRAFPCLSLAYKAYSLETFEVALSNPGJHRKRLVCRASLERVDPRIIRRL
TPDYEPWCKRLVMSGFFRAYIQRDDVELVTAGIDHVEHRIIVDDGLHSDVIVLAT
GPDHAFPRFMQLTGRGIRIDVWODGPHAQTVAIIPGFNPFMMGLGPHSPVGNFPL
TVAESGAHHIVQMIKWRNGEPDTMEPKSAATEAVNTVLRAMPNTVMTGDSWVL
```

Query Match 21.2%; Score 312.6; DB 1; Length 327650;
Best local similarity 54.0%; Pred. No. 2.ee-54;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```
Qy 65 ACTGGTTTGGACCCGACAGCTTCATTTCTTGAATGAAACCTCCCAATATGACAAACC 124
Db 21614 ACGGGGCTGACACGGGGGATCTTTCTGACCTGGGATACGAGCCCTGGCGGTGCCGG 21555
Qy 125 TCAATCAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGAAT 184
```

Db 21554 TAGCCGACTCTATGCGCGGAGCTGGCCGCGCTCGGGCGGCCACCGGCGCGAGCAAT 21495
 QY 185 CGGTATTACTGAGAAAGGSCACCGTGAAGGCAATTCGGTAGCCGTATTGTCGAAAT 244
 Db 21494 CGGTGCAACCGGTGAGGAGCCGTATTGGGGCGCGGGTGGCGGTGGCTGTGAAAT 21435
 QY 245 TTTCTCTCTCGCGGATCTTTTGAGCAGCGTCCGCTGCGTGCATCATGAAGGCGATTC 304
 Db 21434 TCGACTCTCTGGGGCGGTGATTTGGGGTGGCAGCGGCCGAGATCAACCGCGCGCTG 21375
 QY 305 ACCGCGCAGAGAGCTGAAATCTCCACTGCTGTCTCTCCCTGCTTCCGGTGGTGGCGCA 364
 Db 21374 AGCGGGCAGACCGCCAGCGGCTGCCGTACTGCGCTCACCAAGCTCGGAGGACCCGCA 21315
 QY 365 TGCAGAGAGACATGAGCTTTGTTCATGATGTTGCCATTAACCGGCGTGGCAGCGTC 424
 Db 21314 TGCAGAGAGACAGGTCGCGTTCGTGCAAGATGTGAAGTCCGTGGGCGCATTCAGCTGC 21255
 QY 425 ACCCGAGGCGCATTTGCGCTTCTGCTGTATTTGGCAATCCAGATGGGTGGCGCA 484
 Db 21254 ACAACAGGCGCGCTGCGCTTACCTGTCTATTGGCCATCCGACACGSGTGAATTT 21195
 QY 485 TGGCTCTGTTGGGTTTCATCTGGGCATCTCACTTTTGGGAAACCGGCGCGCAATAGTT 544
 Db 21194 TCGCTCTGTTGGGCTCTGCTGGGCGCATCTCACCGTCCGAGCGGGCGCTGATCGCT 21135
 QY 545 TCTTGGGTCTTGGCGTGTGAGTTAACACTGGGCGATCGGCTTCCAGACGTTGTGCAGC 604
 Db 21134 TTTTGGGACACAGGTTCTATGATGTTGCTTATGGGCAACCTTCCATCCGGGCAAA 21075
 QY 605 AGCGGAGATTTGGTGAAACTGGTGATTTGATGAAATTTGTGCGCACTCCAATTGC 664
 Db 21074 CGGCGAAGATCTAGGGCGGCGCATGGGATCATGACGGGCTGTTGCACTGACCGGCTAC 21015
 QY 665 GTGACGCGGTGCAAAACCTCAAGTTATTCAAGCGGTAGAGCAACGATCGTTT 724
 Db 21014 GACCATCTGTGATCGTGTGAGCGTGTGATGACGCTCCGAAACGCTTCCGGCAC 20955
 QY 725 CTGCAACAATCTTGGCGTGGCA---CTTCCGGTATGAGGCGGATTTGCGCTTCTGTC 781
 Db 20954 GCGAGACCGCGCGCCCTTACCCGATGTGCCACGTGGGACTCGGTGTGCGATCGGCC 20895
 QY 782 ACCCGAGAGGCTTGGAAATCGGGAGATTATGAAACGTTGGGGCAGACGTCTCAAGC 841
 Db 20894 GCGCGGACCGGCGGCGGTGACGAGCTACTGCGACACGCGCCACCGACGGGTTGT 20835
 QY 842 TTTCTGTGCGGTGCTGCGCGATTTAGCGCGGCTGTGCGGCTTCCCTGCGCGCATCG 901
 Db 20834 TGTGAGGAACGATCAAGGCGCAAGCG---GCGACACGCTGTGCGCTGGCCGCTTGG 20778
 QY 902 GGGGCGGCGCGGTGTGCTGATTGGGCAAG-----GATGCGCGCTTCAAGC 946
 Db 20777 GCGGCAACCCACGCGTGTCTCTCGGCGCAAAAGGCGAGTAGCGCGGCGGGAACACTG 20718
 QY 947 TTGGGCGCGAGAGCTGCGTTTGTGCGCTGCGCATTTGCTGCGCGCGAGCTAAAC 1006
 Db 20717 TCGGCGCGCGTGTGCGTGAAGGAGCCGACGCGGAGTGGCGCTTCCGCGGAGCTGTGCC 20658
 QY 1007 TGGCATGTGTCAATCATCAACACTTCGCGCGCGAAATTTGTGCGAGCGGCTTGAAGAC 1066
 Db 20657 TGGCGCTGTGTGTCAATTGACGCGGCGGACCGCGTGTGTGCGCGGAGCGAAGAG 20598
 QY 1067 TCGCATGCGAAGTTCGATTGGCGGACCTTGTCCAAGCTTATCGACGCTCCCTCCCA 1126
 Db 20597 GCGGCTGCGCGCGCATTCGCGATTGCTGCGAGCTGCTCAACGCTGATACCCCGA 20538
 QY 1127 CCGTTTCGATCATTTATGATGAGGCGTTTGGCGGTGCGCGCTGCGCATGCTGCCGCG 1186
 Db 20537 CCGTGTCAATCTGTGCGGCGAGGCGAGCGCGCGCGCTGCGGATTTGCGCGCG 20478
 QY 1187 ATTGTGTACGCGCGCAAAACGCGTGTGTTCGCAATTGCGCACAGAGGCGCTCGG 1246
 Db 20477 ACCGGGTGTGCGCGCATTCACGCGGTGTGCGCTGCGCTTCCGGAAGGAGCGAGCG 20418

QY 1247 CCATCTCTTCGCGCGACACCAACGCGCGGAAATCATAGAGGAGCGAGCGTGCAGG 1306
 Db 20417 CGATGTGTTCGAGACACTGTCTCATGCGCGGAACTGTGCGCCCAAGGCAATCGGT 20358
 QY 1307 CGCAGCACTTTAAGCCAAAGGCTTATCGACGGGATGTGCGCGGAAACGAGCACTTTG 1366
 Db 20357 CGGCGGACTTATGAAATCGGGGATTTGTGACACCATGCTGTGCGGAGTATCCCGAGCGCG 20298
 QY 1367 TTGAAGA 1373
 Db 20297 CAGACGA 20291

Search completed: November 13, 2003, 00:10:39
 Job time : 5538.03 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 15:05:16 ; Search time 407.983 Seconds
(without alignments)
9746.164 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtggaagacgtttccgcac.....gattcagcgtttacgcac 1473

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

N.Geneseq.190un03.*
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.*
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1473	100.0	1473	22	AAH65894
2	1473	100.0	1473	22	AAH65894
3	1473	100.0	1473	22	AAH65894
4	1473	100.0	1473	22	AAH65894
5	1473	100.0	1473	22	AAH65894
6	1473	100.0	1473	22	AAH65894
7	1473	100.0	1473	22	AAH65894
8	1473	100.0	1473	22	AAH65894

9	122.8	8.3	873	23	AAH54163	Pseudomonas aeruginosa
10	122.2	8.3	876	25	ABZ39000	N. gonorrhoeae nuc
11	121	8.2	867	23	AAH53191	Enterococcus faecalis
12	121	8.2	867	23	AAH53191	Enterococcus faecalis
13	121	8.2	867	23	AAH53191	Enterococcus faecalis
14	119.2	8.1	876	24	ABH65750	Helicobacter pylori
15	119.2	8.1	876	24	ABH65750	Helicobacter pylori
16	119.2	8.1	876	24	ABH65750	Helicobacter pylori
17	119.2	8.1	876	24	ABH65750	Helicobacter pylori
18	118.6	8.1	876	24	ABH65750	Helicobacter pylori
19	118.6	8.1	876	24	ABH65750	Helicobacter pylori
20	108.6	7.4	933	23	AAH56217	Salmonella typhi
21	108	7.3	993	23	AAH56217	Salmonella typhi
22	102.8	7.0	915	23	AAH56217	Salmonella typhi
23	99.8	6.8	318	23	AAH88487	Streptococcus pneumoniae
24	99.8	6.8	318	23	AAH88487	Streptococcus pneumoniae
25	97.8	6.6	867	23	AAH55883	Streptococcus pneumoniae
26	96.8	6.6	867	23	AAH55883	Streptococcus pneumoniae
27	96.2	6.5	864	25	ABX06060	Streptococcus pneumoniae
28	96.2	6.5	867	22	AAH90763	Streptococcus pneumoniae
29	96.2	6.5	891	23	AAH53478	Streptococcus pneumoniae
30	96.2	6.5	1595	19	AAH296348	Streptococcus pneumoniae
31	96.2	6.5	19702	19	AAH52140	Streptococcus pneumoniae
32	96.2	6.5	1830121	17	AAH42063	Streptococcus pneumoniae
33	96.2	6.5	2162598	25	ABH56454	Streptococcus pneumoniae
34	94.2	6.4	873	24	ABH66650	Streptococcus pneumoniae
35	94.2	6.4	2155561	24	ABH71527	Streptococcus pneumoniae
36	93.6	6.4	495269	24	ABH67195	Streptococcus pneumoniae
37	93.6	6.4	3011208	24	ABH69245	Streptococcus pneumoniae
38	93	6.3	867	23	AAH55630	Streptococcus pneumoniae
39	93	6.3	891	22	AAH90875	Streptococcus pneumoniae
40	92.8	6.3	1001	22	AAH91440	Streptococcus pneumoniae
41	92.8	6.3	1001	24	ABH37820	Streptococcus pneumoniae
42	91	6.2	2365589	24	ABH30521	Streptococcus pneumoniae
43	87.4	6.0	2944528	24	ABH3041	Streptococcus pneumoniae
44	87.4	5.9	954	18	AAH90220	Streptococcus pneumoniae
45	87.4	5.9	28626	22	AAH28528	Streptococcus pneumoniae

ALIGNMENTS

RESULT 1	AAH65894	standard; DNA, 1473 BP.
ID	AAH65894	
AC	AAH65894	
XX		
DT	26-SEP-2001	(first entry)
DE	C glutamicum coding sequence fragment SEQ ID NO: 929.	
XX		
XX	Corynebacterium; amino acid synthesis; vitamin; saccharide;	
KW	organic acid synthesis; ds.	
XX		
OS	Corynebacterium glutamicum.	
XX		
PN	EP1108790-A2.	
XX		
PD	20-JUN-2001.	
XX		
PF	18-DEC-2000; 2000EP-0127688.	
XX		
PR	16-DEC-1999; 99JP-0377484.	
XX		
PR	07-APR-2000; 2000JP-0159162.	
XX		
PR	03-AUG-2000; 2000JP-0280988.	
XX		
PA	(KYOW) KYOWA HAKKO KOGYO KK.	
XX		
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;	
XX		
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;	
XX		
DR	WPI; 2001-376931/40.	

XX 29-NOV-2000.
PD
XX
XX 10-MAY-2000; 2000EP-0109842.
PF
XX
XX 27-MAY-1999; 99DE-1024365.
PR
XX
XX (DEGS) DEGUSSA-HUELS AG.
PA (KERU) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
XX Tlig Y, Eggeling L, Elkmanns B, Sahn H, Moeckel B;
PI WPI; 2001-042411/06.
XX
XX DR P-PSDB; AAB45789.
XX
XX
XX Cloned Corynebacterium glutamicum accda gene useful for producing
PT transformed coryneform bacteria producing increased yields of L-amino
XX acids, especially L-lysine -
XX
XX
XX Disclosure; Page 12-15; 20pp; German.
PS
XX
XX This invention describes a novel cloned Corynebacterium glutamicum DNA
CC (1), replicable in coryneform microorganisms coding for an accda gene.
CC The invention also describes (1) a polypeptide derived from (1) having
CC a fully defined 491 aa sequence; (2) coryneform microorganisms
CC transformed with one or more copies of (1); (3) the shuttle vector
CC pZlacaccda contained in Corynebacterium glutamicum DSM 12785; and
CC (4) a process for producing L-amino acids, comprising culturing a
CC coryneform bacterium that overexpresses the accda gene. Coryneform
CC bacteria transformed with (1) so that they overexpress the accda gene are
CC useful for producing L-amino acids, especially L-lysine, which are
CC useful in animal feeds, in human medicine and in the pharmaceutical
CC industry.

SQ Sequence 1473 BP; 275 A; 414 C; 456 G; 328 T; 0 other;

Query Match	100.0%	Score 1473	DB 22	Length 1473
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1473	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	GTGGGAAGACGGTTTTCGAGCTATGGGTGTGGGGCATGGAACAACCTTCAGATTGACGCTC	60
Db	1	GTGGGAAGACGGTTTTCGAGCTATGGGTGTGGGGCATGGAACAACCTTCAGATTGACGCTC	60
QY	61	ATGAGCTCGGTTTGGACCTTGACAGCTTCATTCTTGGAAATGAAACTCCCAATAATGAC	120
Db	61	ATGAGCTCGGTTTGGACCTTGACAGCTTCATTCTTGGAAATGAAACTCCCAATAATGAC	120
QY	121	AACCCATCAAGGCTATGSCAAGACCTTGGAGCGGGGCTCGAAGCAAGGCCAAATCCGAT	180
Db	121	AACCTCAATCAAGGCTATGSCAAGACCTTGGAGCGGGGCTCGAAGCAAGGCCAAATCCGAT	180
QY	181	GAATCGGTAAATTACTGGAAGAGGCAACCGTGAAGGGCATTCGGTAGCCGTTAATTTGTCC	240
Db	181	GAATCGGTAAATTACTGGAAGAGGCAACCGTGAAGGGCATTCGGTAGCCGTTAATTTGTCC	240
QY	241	GATTTTTCCTTCCTCGGCGGTTCTTTGGSCAAGGTCGTCGTCGTGGCATCATGAAGGGG	300
Db	241	GATTTTTCCTTCCTCGGCGGTTCTTTGGSCAAGGTCGTCGTCGTGGCATCATGAAGGGG	300
QY	301	ATTCAACCGGCCACAGAGCTGAACTCCACATGCTGATCTCCCTCTCCGATGGTGGG	360
Db	301	ATTCAACCGGCCACAGAGCTGAACTCCACATGCTGATCTCCCTCTCCGATGGTGGG	360
QY	361	CGCATGCAAGAAACAATCGAGCTTTTGTCAATGATGATGTCATTAACCGCGGCTGTGGAG	420
Db	361	CGCATGCAAGAAACAATCGAGCTTTTGTCAATGATGATGTCATTAACCGCGGCTGTGGAG	420
QY	421	CGTCAACCGCGAGGCGCATTTGGCGTTCTGATGTAATTTCCGCAATCCACAGATGGGTGAC	480
Db	421	CGTCAACCGCGAGGCGCATTTGGCGTTCTGATGTAATTTCCGCAATCCACAGATGGGTGAC	480
QY	481	GCCATGCGCTCGTGGGGTTTCATCTGGGCACTTCACTTTTGGGAACCGGCGCGCAGATA	540

Db	441	GCATGCGCTCGTGGGGTTCACTTGGGCATCTTATTCGGAAACCCGGCGGCAGATA	540
Qy	541	GGTTTCTCTGGGTCCTCGCGGTGGAGTTAACCACTGGGCATGCGCTTCAGACGGTGTG	600
Db	541	GGTTTCTCTGGGTCCTCGCGGTGGAGTTAACCACTGGGCATGCGCTTCAGACGGTGTG	600
Qy	601	CAGCAGCGCGAGAAATTGGTGAATACTGGTGTGATTGATGAAATTGTGTGCCACTTCAA	660
Db	601	CAGCAGCGCGAGAAATTGGTGAATACTGGTGTGATTGATGAAATTGTGTGCCACTTCAA	660
Qy	661	TTGGGTGACGGGTGGGCAAAAACCTTCAAGTTTATTACGCCGGTAAAGGCACAGATCGT	720
Db	661	TTGGGTGACGGGTGGGCAAAAACCTTCAAGTTTATTACGCCGGTAAAGGCACAGATCGT	720
Qy	721	TTTTTCTCAACAACCTCTGCGCGTGGCACTTCGGGTGATGAGAGCCGATTGCGCTTCGT	780
Db	721	TTTTTCTCAACAACCTCTGCGCGTGGCACTTCGGGTGATGAGAGCCGATTGCGCTTCGT	780
Qy	781	GACCCGACAGAGGCTTGAAATCGGGAGATTATGAAACGTTTGGGGGCAGACGTGTCAG	840
Db	781	GACCCGACAGAGGCTTGAAATCGGGAGATTATGAAACGTTTGGGGGCAGACGTGTCAG	840
Qy	841	CTTTCCTGGTGGCGCGTCTGCGGCATTGAGCCCGGCTGTGCGGCTTGCCCTGGCGGCATC	900
Db	841	CTTTCCTGGTGGCGCGTCTGCGGCATTGAGCCCGGCTGTGCGGCTTGCCCTGGCGGCATC	900
Qy	901	GGGGGCGCGGCGGTGTCGTGATTGGGACAGATTCGCGCTTACGCTTGAGGCGCAGAG	960
Db	901	GGGGGCGCGGCGGTGTCGTGATTGGGACAGATTCGCGCTTACGCTTGAGGCGCAGAG	960
Qy	961	CTGGGTTTTCGCGCGTCTGTGGAATTCGTGTCGCGCGCAGACTAACTGCGCATCTGTGTC	1020
Db	961	CTGGGTTTTCGCGCGTCTGTGGAATTCGTGTCGCGCGCAGACTAACTGCGCATCTGTGTC	1020
Qy	1021	ATCATTCGACACTTCGCGCGCCGGAATTTGTGCGACGGCGCTGAGAGAGTCTGGCATTCGCAAGC	1080
Db	1021	ATCATTCGACACTTCGCGCGCCGGAATTTGTGCGACGGCGCTGAGAGAGTCTGGCATTCGCAAGC	1080
Qy	1081	TCGATTGGGCGCACCTTGTCCAAAGCTTATTCGACGCTCCCTCCACGCTTTCGGTCATT	1140
Db	1081	TCGATTGGGCGCACCTTGTCCAAAGCTTATTCGACGCTCCCTCCACGCTTTCGGTCATT	1140
Qy	1141	ATTGGTCAGGGCGTTGGCGGTGCGCGCTGCGCATGCTGCCCGCATCTGCTTCAACGCG	1200
Db	1141	ATTGGTCAGGGCGTTGGCGGTGCGCGCTGCGCATGCTGCCCGCATCTGCTTCAACGCG	1200
Qy	1201	GCCGAAAAACGCGTGGCTGTCCGCAATTGCGACACAGAGGCGCGCTCGGCACTCTTCCGCG	1260
Db	1201	GCCGAAAAACGCGTGGCTGTCCGCAATTGCGACACAGAGGCGCGCTCGGCACTCTTCCGCG	1260
Qy	1261	GACACCAACACGCGCGGGAATAATCAAGGACAAAGGCGTGCAGAGCGACGACCTTTTA	1320
Db	1261	GACACCAACACGCGCGGGAATAATCAAGGACAAAGGCGTGCAGAGCGACGACCTTTTA	1320
Qy	1321	AGCCAAAGGCGCTTATTCGACGGAGATGTCGCGCAAAACGAGCACTTGTGTTGAAGAAATTCTC	1380
Db	1321	AGCCAAAGGCGCTTATTCGACGGAGATGTCGCGCAAAACGAGCACTTGTGTTGAAGAAATTCTC	1380
Qy	1381	GGCAACAATCAGCAACGCCCTTCCGAATTGGATTAACAATCCGGAAGAGGCGGAGCGCGAC	1440
Db	1381	GGCAACAATCAGCAACGCCCTTCCGAATTGGATTAACAATCCGGAAGAGGCGGAGCGCGAC	1440
Qy	1441	AGTGGCTTCAACAAGATTGAGCGTTTGAAGCGAG	1473
Db	1441	AGTGGCTTCAACAAGATTGAGCGTTTGAAGCGAG	1473
RESULT 3			
AAAF67755			
1D AAF67755 standard; DNA; 1575 BP.			
XX AAF67755;			

```
RESULT 3
AAAF67755
ID   AAAF67755 standard; DNA; 1575 BP
XX
AC   AAAF67755;
```

XX 11-APR-2001 (first entry)
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:25.
XX
XX Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
XX membrane construction and membrane transport protein; petroleum spill;
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
OS Corynebacterium glutamicum.
XX
XX WO200100805-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-1B00926.
XX
XX 25-JUN-1999; 99US-0141031.
XX 08-JUL-1999; 99DE-1031454.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031563.
XX 09-JUL-1999; 99DE-1032122.
XX 09-JUL-1999; 99DE-1032124.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032128.
XX 09-JUL-1999; 99DE-1032180.
XX 09-JUL-1999; 99DE-1032182.
XX 09-JUL-1999; 99DE-1032190.
XX 09-JUL-1999; 99DE-1032191.
XX 09-JUL-1999; 99DE-1032209.
XX 09-JUL-1999; 99DE-1032212.
XX 09-JUL-1999; 99DE-1032227.
XX 09-JUL-1999; 99DE-1032228.
XX 09-JUL-1999; 99DE-1032229.
XX 09-JUL-1999; 99DE-1032230.
XX 14-JUL-1999; 99DE-1032927.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 27-AUG-1999; 99DE-1040764.
XX 27-AUG-1999; 99DE-1040765.
XX 27-AUG-1999; 99DE-1040766.
XX 27-AUG-1999; 99DE-1040830.
XX 27-AUG-1999; 99DE-1040831.
XX 27-AUG-1999; 99DE-1040832.
XX 27-AUG-1999; 99DE-1040833.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041395.
XX 03-SEP-1999; 99DE-1042077.
XX 03-SEP-1999; 99DE-1042078.
XX 03-SEP-1999; 99DE-1042079.
XX 03-SEP-1999; 99DE-1042088.
XX
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-071486/08.
XX P-PSDB; AAB76522.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
XX identifying C. glutamicum or related bacteria, and as markers for
XX transformation -
XX
XX Claim 3; Page 173-175, 119pp; English.
XX
XX AAB7743 to AAB68080 encode the Corynebacterium glutamicum membrane
XX construction and membrane transport (MCT) proteins given in AAB76510 to
XX AAB76847. The MCT nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine

CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAB68082 and AAB68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
XX SQ Sequence 1575 BP; 309 A; 434 C; 474 G; 358 T; 0 other;
Query Match 100.0%; Score 1473; DB 22; Length 1575;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGAGAAAGCTTTTCCGACATATGATGGGCGATGAAACACACTTGACATTCACCTC 60
DB 80 GTGGAGAAAGCTTTTCCGACATATGATGGGCGATGAAACACACTTGACATTCACCTC 139
QY 61 ATAGACTGGTTTGGAGACCTTGACAGCTTCATTTCTTGAATGAATCCCAATATGAC 120
DB 140 ATAGACTGGTTTGGAGACCTTGACAGCTTCATTTCTTGAATGAATCCCAATATGAC 199
QY 121 AACCTCAATCAAGCTATGACAGACCTTGAGAGGGGCTGAAACAGGCCAATGAGAT 180
DB 200 AACCTCAATCAAGCTATGACAGACCTTGAGAGGGGCTGAAACAGGCCAATGAGAT 259
QY 181 GAATCGGTAATTAATGAGAGGACCGTGGAGGGCAATCCGTAAGCGTTATTTGTC 240
DB 260 GAATCGGTAATTAATGAGAGGACCGTGGAGGGCAATCCGTAAGCGTTATTTGTC 319
QY 241 GAATTTTCTTCTCGGCGGTTCTTTGGGACGAGTGCCTGCTGTCGCATCATGAAGCG 300
DB 320 GAATTTTCTTCTCGGCGGTTCTTTGGGACGAGTGCCTGCTGTCGCATCATGAAGCG 379
QY 301 ATTACCGCGCCGACAGAGCTGAACTCCACTGCTGCTCCCTCCGCTCCGCGTGGCG 360
DB 380 ATTACCGCGCCGACAGAGCTGAACTCCACTGCTGCTCCCTCCGCTCCGCGTGGCG 439
QY 361 CGCATGCGAGAAACAATCGAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 420
DB 440 CGCATGCGAGAAACAATCGAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 499
QY 421 CGTACCGCGAGGCGCATTTGCTCCGTTCTGATGATTTGGCAATCCACGATGGTGGC 480
DB 500 CGTACCGCGAGGCGCATTTGCTCCGTTCTGATGATTTGGCAATCCACGATGGTGGC 559
QY 481 GCCATGGCTGCTGGGGTTCATCTGGGCATCTCATCTTTGGGAAACCGGCGCCAGATA 540
DB 560 GCCATGGCTGCTGGGGTTCATCTGGGCATCTCATCTTTGGGAAACCGGCGCCAGATA 619
QY 541 GATTTCTGAGGCTCGGCTGAGGATTAACAATGAGGCAATGCGCTTCCAGACGATGTC 600
DB 620 GATTTCTGAGGCTCGGCTGAGGATTAACAATGAGGCAATGCGCTTCCAGACGATGTC 679
QY 601 CAGCAGGCGAGAAATTTGTAATAAATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 680 CAGCAGGCGAGAAATTTGTAATAAATGATGATGATGATGATGATGATGATGATGATGAT 739
QY 661 TTGCGTGAAGCGGTGGCAAAAACCTCAAGTTATTAAGCGGTGAGAGGCAACGATTCGT 720
DB 740 TTGCGTGAAGCGGTGGCAAAAACCTCAAGTTATTAAGCGGTGAGAGGCAACGATTCGT 799
QY 721 TTTTCTCAACAACCTCGGCTGAGCACTTCGCTGATGAGAGGCAATGCGCTTCTCGT 780
DB 800 TTTTCTCAACAACCTCGGCTGAGCACTTCGCTGATGAGAGGCAATGCGCTTCTCGT 859
QY 781 GACCGCAGAGCGCTGGAATCGGGAATTAATGAAAAGTTGGGGGAGAGAGTGTCAAG 840
DB 860 GACCGCAGAGCGCTGGAATCGGGAATTAATGAAAAGTTGGGGGAGAGAGTGTCAAG 919
QY 841 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 920 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 979

[illegible][illegible]

```
Db      1228 TTTTCTCCACAACTCTCGGCGTGGCACTTCCGGATGAGGCGGATTTGCCGTTCTCTGT 1287
Qy      781  GACCCGAGAGGCTCGAATCGGAGATTATGAAAAGTTGGGCGAGACGTCGTCAAG 840
Db      1288 GACCCGAGAGGCTCGAATCGGAGATTATGAAAAGTTGGGCGAGACGTCGTCAAG 1347
Qy      841  CTTTCTGCTGCGCTGCTGGCGCATTTAGAGCCGCGTGGCGGCTTCCCTTGGCCCGCATC 900
Db      1348 CTTTCTGCTGCGCTGCTGGCGCATTTAGAGCCGCGTGGCGGCTTCCCTTGGCCCGCATC 1407
Qy      901  GGGGGCGCGCCGCTGCTGCTGATTTGGGCGAGATGCGCGCTTCAAGCTTGGCCCGCAGAG 960
Db      1408 GGGGGCGCGCCGCTGCTGCTGATTTGGGCGAGATGCGCGCTTCAAGCTTGGCCCGCAGAG 1467
Qy      961  CTGCGTTTGGCGCTGCTGCGATTTGCTGGCGCGGCGAGCTTAAACCTGCGCATGTGTCC 1020
Db      1468 CTGCGTTTGGCGCTGCTGCGATTTGCTGGCGCGGCGAGCTTAAACCTGCGCATGTGTCC 1527
Qy      1021 ATCAATCGACACTTCCGCGCGGCAATTTGCGAGGCGGCTGAGAGCTTGGGATGCGAAGC 1080
Db      1528 ATCAATCGACACTTCCGCGCGGCAATTTGCGAGGCGGCTGAGAGCTTGGGATGCGAAGC 1587
Qy      1081 TCGATTTGGCGGACCTTGTCCAAAGCTTATCGACGCTCCCTCCCAACGTTTGGGTCAAT 1140
Db      1588 TCGATTTGGCGGACCTTGTCCAAAGCTTATCGACGCTCCCTCCCAACGTTTGGGTCAAT 1647
Qy      1141 ATTGCTCAGGCGCTTGGCGGCTGCGCGCTGCGCATGCTCCCGCATGCTGCTCAAGC 1200
Db      1648 ATTGCTCAGGCGCTTGGCGGCTGCGCGCTGCGCATGCTCCCGCATGCTGCTCAAGC 1707
Qy      1201 GCCGAAAAGCGGTGGCTGTCCGCAATTTGCACAGAGGCGGCTGGGCACTCTCTCCGC 1260
Db      1708 GCCGAAAAGCGGTGGCTGTCCGCAATTTGCACAGAGGCGGCTGGGCACTCTCTCCGC 1767
Qy      1261 GACACCAACACAGCGCGCGGAATATCATAGAGCGACAGGCGTGCAGCGCACTTTTA 1320
Db      1768 GACACCAACACAGCGCGCGGAATATCATAGAGCGACAGGCGGCGCACTTTTA 1827
Qy      1321 AGCCAAAGGCTTATTCAGCGGATGCTGCCCGAAGCCGAGCACTTTGTGAAGAAATCTC 1380
Db      1828 AGCCAAAGGCTTATTCAGCGGATGCTGCCCGAAGCCGAGCACTTTGTGAAGAAATCTC 1887
Qy      1381 GGCACAATCAGCAAGCCCTCTCCGAATTTGATACATTCGAGAGAGGCGGAGACGGAC 1440
Db      1888 GGCACAATCAGCAAGCCCTCTCCGAATTTGATACATTCGAGAGAGGCGGAGACGGAC 1947
Qy      1441 AGTCGCTTCAACGATTTGAGCGTTAGCGAG 1473
Db      1948 AGTCGCTTCAACGATTTGAGCGTTAGCGAG 1980

RESULT 5
AAH68526/c
ID      AAH68526 standard; DNA; 349980 BP.
XX      AAH68526;
AC      AAH68526;
XX      26-SEP-2001 (first entry)
XX      C glutamicum coding sequence fragment SEQ ID NO: 7061.
XX      Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX      organic acid synthesis; ds.
XX      Corynebacterium glutamicum.
OS      Corynebacterium glutamicum.
PN      EPI108790-A2.
XX      20-JUN-2001.
XX      18-DEC-2000; 2000EP-0127688.
XX
```

```
PR      16-DEC-1999; 99JP-0377484.
PR      07-APR-2000; 2000JP-0159162.
PR      03-AUG-2000; 2000JP-0280988.
XX
XX      (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI      Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX      WPI; 2001-376931/40.
XX
PT      Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT      mutation point of a gene, measuring expression of a gene, analysing
PT      expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX      Disclosure; SEQ ID NO: 7061; 246bp + Sequence Listing; English.
XX
XX      The present invention provides a number of nucleotide and protein
XX      sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
XX      are useful for identifying the mutation point of a gene derived from a
XX      mutant of coryneform bacterium, measuring expression amount and
XX      analysing the expression profile or expression pattern of a gene derived
XX      from Corynebacterium bacterium, and identifying a homologue of a gene derived
XX      from coryneform bacterium. Corynebacterium bacteria are useful for producing
XX      amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX      particularly L-lysine. The present sequence is a nucleic acid described
XX      in the exemplification of the invention.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from the
XX      European Patent Office.
XX
SQ      Sequence 349980 BP; 79274 A; 90638 C; 98727 G; 81341 T; 0 other;

Query Match      100.0%; Score 1473; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GTGAGAAAGGTTTCCGACTATGTTGTGGGCAATGGAACACTTGACATTTGACGCTC 60
Db      281114 GTGAGAAAGGTTTCCGACTATGTTGTGGGCAATGGAACACTTGACATTTGACGCTC 281055
Qy      61 ATAGACTCGGTTTGAACCTGACAGCTTATTTCTTGAATGAATCCCAATATGAC 120
Db      281054 ATAGACTCGGTTTGAACCTGACAGCTTATTTCTTGAATGAATCCCAATATGAC 280995
Qy      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGGCCAAATGCGAT 180
Db      280994 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGGCCAAATGCGAT 280935
Qy      181 GAATCGGTAATTAATGAGAAAGGACCGTGAAGGCAATTCGGTAGCGTATTTGTCC 240
Db      280934 GAATCGGTAATTAATGAGAAAGGACCGTGAAGGCAATTCGGTAGCGTATTTGTCC 280875
Qy      241 GATTTTCTTCTCTCGCGCGGTTCTTTGGGACCGTCCGCTGCGTGCATCATGAAGCGC 300
Db      280874 GATTTTCTTCTCTCGCGCGGTTCTTTGGGACCGTCCGCTGCGTGCATCATGAAGCGC 280815
Qy      301 ATTCAACCGCGCACAGAGCTGAAATCTCCACTGCTGTCTTCCCTGTTCCGTTGTCG 360
Db      280814 ATTCAACCGCGCACAGAGCTGAAATCTCCACTGCTGTCTTCCCTGTTCCGTTGTCG 280755
Qy      361 CGCATGACAGAAAGCAATTCAGCTTTTGTATGATGATGTCGCCAATACCGGCGTGTGAG 420
Db      280754 CGCATGACAGAAAGCAATTCAGCTTTTGTATGATGATGTCGCCAATACCGGCGTGTGAG 280695
Qy      421 CGTCACGCGGAGGCGCATTTTCCGCTTCTGATTTTGGCAATCCACGATGGGTGAC 480
Db      280694 CGTCACGCGGAGGCGCATTTTCCGCTTCTGATTTTGGCAATCCACGATGGGTGAC 280635
Qy      481 GCCATGGGCTGTGGGGTTCACTGCGCATCTCACTTTTGGGAACCGGCGCGCAGATA 540
Db      280634 GCCATGGGCTGTGGGGTTCACTGCGCATCTCACTTTTGGGAACCGGCGCGCAGATA 280575
```

```

QY 541 GGTTCCTGAGTCTGCGTGTGAGTTAAACCACTGGCATCGCTTCCAGACGGTGTG 600
DB 280574 GGTTCCTGAGTCTGCGTGTGAGTTAAACCACTGGCATCGCTTCCAGACGGTGTG 280515
QY 601 CAGCAGCGGAGAAATTTGGTGAATACTGGTGTGATGAAATTTGTGTGCCATCCAA 660
DB 280514 CAGCAGCGGAGAAATTTGGTGAATACTGGTGTGATGAAATTTGTGTGCCATCCAA 280455
QY 661 TTGCGTGACGGGTGGCAAAACCTCAAGTATTACCGCTTAAAGCAACGATTCGT 720
DB 280454 TTGCGTGACGGGTGGCAAAACCTCAAGTATTACCGCTTAAAGCAACGATTCGT 280395
QY 721 TTTTCTCAACAACCTCTGCGTGTGACCTTCCGATGAGAGCGATTGCGCGTTCGT 780
DB 280394 TTTTCTCAACAACCTCTGCGTGTGACCTTCCGATGAGAGCGATTGCGCGTTCGT 280335
QY 781 GACCCGACAGAGCTTGAATTCGGGAGATTATGAAACGTTGGGGCAGACGTCGTAAG 840
DB 280334 GACCCGACAGAGCTTGAATTCGGGAGATTATGAAACGTTGGGGCAGACGTCGTAAG 280275
QY 841 CTTTCTGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 280274 CTTTCTGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280215
QY 901 GGGGGCCGGCCCGTGTGCTGATTTGGGACAGATTCGCGCTTCAAGCTTGGGCGCAGAG 960
DB 280214 GGGGGCCGGCCCGTGTGCTGATTTGGGACAGATTCGCGCTTCAAGCTTGGGCGCAGAG 280155
QY 961 CTGCGTTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 280154 CTGCGTTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280095
QY 1021 ATCATTCGACACTTCGGGCGCGCAATTTGTGTGACGCGGCTGAGAGCTTCGCAATCCGAC 1080
DB 280094 ATCATTCGACACTTCGGGCGCGCAATTTGTGTGACGCGGCTGAGAGCTTCGCAATCCGAC 280035
QY 1081 TCGATTGCGCGACCTTGTGCAAGCTTATGACGCTCCCTCCCAACGCTTTCGGTCAAT 1140
DB 280034 TCGATTGCGCGACCTTGTGCAAGCTTATGACGCTCCCTCCCAACGCTTTCGGTCAAT 279975
QY 1141 ATTGTCAGAGGGGCTTGGCGGCTGCGCGCTGCGCATGCTGCGCGCATGCTGCTAAGCG 1200
DB 279974 ATTGTCAGAGGGGCTTGGCGGCTGCGCGCTGCGCATGCTGCGCGCATGCTGCTAAGCG 279915
QY 1201 GCCGAAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 279914 GCCGAAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 279855
QY 1261 GACACCAACCAACGCGCGGCAATTCATAGAGCAAGAGGCGCTGCGCATCTCTTCCGC 1320
DB 279854 GACACCAACCAACGCGCGGCAATTCATAGAGCAAGAGGCGCTGCGCATCTCTTCCGC 279795
QY 1321 AGCCAAAGGCTTATGACGCGGATTCGCGCAAAACGAGCACTTTGTGGAATAATCTC 1380
DB 279794 AGCCAAAGGCTTATGACGCGGATTCGCGCAAAACGAGCACTTTGTGGAATAATCTC 279735
QY 1381 GGCACAAATCAGAAAGCCCTCTCCGAATTTGATCAATCCGAGAGGCGGAGCGCAAC 1440
DB 279734 GGCACAAATCAGAAAGCCCTCTCCGAATTTGATCAATCCGAGAGGCGGAGCGCAAC 279675
QY 1441 AGTCGCTTCAACGATTTGAGGCTTTAGGCGAG 1473
DB 279674 AGTCGCTTCAACGATTTGAGGCTTTAGGCGAG 279642

```

```

RESULT 6
AAH51981
ID AAH51981 standard; DNA; 1488 BP.
AC AAH51981;
XX
DT 04-SEP-2001 (first entry)

```

```

XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 35.
XX KM Drug target; growth; organism viability; characterisation; ds.
XX OS Mycobacterium tuberculosis.
XX PN W0200135317-A1.
XX PD 17-MAY-2001.
XX PF 13-NOV-2000; 2000WO-US31152.
XX PR 12-NOV-1999; 99US-0165086.
XX PR 12-NOV-1999; 99US-0165124.
XX PR 01-FEB-2000; 2000US-0179531.
XX (REGC ) UNIV CALIFORNIA.
XX PA Eisenberg D, Roststein SH, Marcotte EM;
XX DR WPI; 2001-329193/34.
XX PI P-PSDB; AAG81130.
XX PT Identifying nucleotide or polypeptide sequence for use as drug target,
XX PT involves providing algorithm that analyzes a functional relationship
XX PT between nucleotide or polypeptide sequences, and comparing the
XX PS sequences.
XX PS Disclosure; Page 74; 207p; English.
XX CC This invention relates to a method for identifying a nucleotide or
XX CC polypeptide sequence that may be a drug target, or essential for growth
XX CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
XX CC tuberculosis proteins which are potential drug targets. The DNA and
XX CC protein sequences are used to illustrate the method of the invention. The
XX CC method involves providing an unknown nucleotide or polypeptide sequences,
XX CC and comparing it to a number of sequences along with at least one
XX CC algorithm capable of analyzing a functional relationship between
XX CC nucleotide and polypeptide sequences. The method is useful for
XX CC characterizing the function of nucleic acids and polypeptides that may be
XX CC useful as a target for a drug or essential for the growth or viability of
XX CC an organism.
XX SQ Sequence 1488 BP; 220 A; 509 C; 524 G; 235 T; 0 other;
XX
Query Match 21.2%; Score 312.6; DB 22; Length 1488;
Best Local Similarity 54.0%; Pred. No. 1.5e-78;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
QY 65 ACTCGTTTGGACCTTGACAGCTTCAATTTTGGATGAATACTCCCAATATGACAAAC 124
DB 32 ACGCGGTCTAGACCGGGGATCTTTCGACGTGGATAGCGAGCGCTGGCGGTGCCGG 91
QY 125 TCAATCAAGGCTTATGACAGAGCTTGGAGCGGCTTCGAAAGAGCCCAATGCCATGAT 184
DB 92 TAGCGACTCTTATGCGGGAGAGCTGGCGCGCTGCGGCGCAACCGCGCGAGCGAT 151
QY 185 CGGTAACTACTGGAAGAGCAACCGTGAAGGCAATTCGCTGAGCGCTTATTTTGTCCGATT 244
DB 152 CGGTGACAGACCGGTGAGGAGAGCGGTATTCGAGCGGCGGAGTGGCTGTGAGT 211
QY 245 TTTCTCTCTGCGCGGTTCTTTGGAGCAGTCCGCTGCTGCTGCGATCAATGAGCGGATTC 304
DB 212 TCGACTTCTGCGCGGCTCGATTGAGGAGGCGAGCGCGCAACGAGTCAACCGCGCGTGC 271
QY 305 ACCGCGCACAGAGCTGAATCTCCCATGCTGCTGCTTCCCTGCTTCCGCTGTGCGGCA 364
DB 272 AGCGGCGACCGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331
QY 365 TGCAGAAAGCAATCGAGCTTTTGTGATGATGTGTCCATAACGCGGCTGTGACGCTC 424

```

Db 332 TCGAAGAGGACGCGCTGCTTTCTGCGATGATGAGATGCTGCGGCATCCAGCTGC 391
 Qy 425 ACCGAGAGCGCATTTGCGGTTCTGATGATTTGCGCAATCCACGATGGGTGGCGCA 484
 Db 392 ACAACGAGGCGCGCTGCTTACCTGCTATTTGGCGCATCCGACGAGGTGAGTTT 451
 Qy 485 TGGCTCTGTTGGGTTCTATCTGGGATCTTCTTTTGGCGAACCGCGCGCGAGATAGTT 544
 Db 452 TCGGCTCTGGGCTGCTGGGCGATCTCACGCTGCGACCGCGCGCGCTGATGGCT 511
 Qy 545 TCTGGGCTCTGCGGTGGATGATTAACCACTGGGCGATGCGCTTCCAGACGGTGTGACG 604
 Db 512 TTCTGGACCAACGGGTCTATGATGTTGCTTATGGGACCCCTTCCATCCGCGCTCAA 571
 Qy 605 AGGCGAGATTTGGTGAATCTGTTGATGATGATGATGATGATGATGATGATGATGAT 664
 Db 572 CCGCGAGATCTACGCGCGCATGGGATCATGACGCGGTGCTGCTGATGACCGGCTAC 631
 Qy 665 GTGCAAGCGGTGGCAAAACCTTCAAGGTTATTCAGCGCGTAGAGCAAGATGTTT 724
 Db 632 GACCGATCTGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 691
 Qy 725 CTCCAAACAATCTGCGGTGGCA---CTTCCGATGATGATGATGATGATGATGATGAT 781
 Db 692 CGCAGACCGCGCGCTTACCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
 Qy 782 ACCGCGAGAGCTGGAATCGGGAGATTAAGAACTTTGGGGGAGAGCTGCTGCAAGC 841
 Db 752 GCGCGAGCGCGCGCGCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
 Qy 842 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
 Db 812 TGTCAAGAAAGCATTAAGAGCGAAGCG---GCGACCAAGCTGCTGCTGCTGCTGCT 868
 Qy 902 GGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
 Db 869 GCGGCAACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928
 Qy 947 TTTGGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1006
 Db 929 TCGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988
 Qy 1007 TCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1066
 Db 989 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048
 Qy 1067 TCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1126
 Db 1049 GCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
 Qy 1127 CCGTTTGGCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
 Db 1109 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1168
 Qy 1187 ATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1246
 Db 1169 ACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
 Qy 1247 CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
 Db 1229 CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1288
 Qy 1307 GCGACGCTTTTAAAGCAAGGCTTATGACGGGATGCTGCTGCTGCTGCTGCTGCTGCT 1366
 Db 1289 CCGCGACCTACTGAATGCGGGATGCTGACACCATGCTGCGGAGTACCCCGAGCGCG 1348
 Qy 1367 TTGAGA 1373
 Db 1349 CAGACGA 1355

RESULT 7
 AA199683/c

ID AA199683 standard; DNA: 4403765 BP.
 XX
 AC AA199683;
 AC
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 XX
 KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
 OS
 XX Mycobacterium tuberculosis.
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 DR WPI: 2001-647261/74.
 XX
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 PT 1551 and H37Rv differ -
 XX
 PS Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
 XX
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
 CC
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
 SQ
 Query Match 21.2%; Score 312.6; DB 22; Length 4403765;
 Best Local Similarity 54.0%; Pred. No. 5.4e-77; Indels 21; Gaps 3;
 Matches 717; Conservative 0; Mismatches 589;
 Qy 65 ACTGCTTTGAGACCTGACAGCTTATTTTGGAAATGAAATCCCAATATGACAAAC 124
 Db 1008148 ACGGCTGCTAGACCGGGGATCTTCTGCACTGAGATGAGACCGCTGCGGTGCGG 1008089
 Qy 125 TCAATCAAGGCTATGACAGACCTTTGAGCGGCTCGAAGACCAAGCCAAATGCAAT 184
 Db 1008088 TAGCGCACTCTATGCGGGAGGTGGCGCGCTCGGGCGCAACCGCGGAGCAAT 1008029
 Qy 185 CGGTAATTAATGAGAAAGGACCGTGAAGGCAATTCGGGTAGCCGTTATTTGTCCGATT 244
 Db 1008028 CCGTCAAGCCGGTAGAGACCGGTAATTCGGGCGGCGGTGCGCTGTGCTGTAGT 1007969
 Qy 245 TTCTCTCTGCGCGCTTCTTTGGGACCGTTCGGGTGCTGCTGCTGCTGCTGCTGCTGCT 304
 Db 1007968 TCGACTTCTGCGCGGCTCGCATTTGGGATGAGAGGCGGCAAGATCAACCGCGCGT 1007909
 Qy 305 ACCGCGCAAGAGTGAATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 364
 Db 1007908 AGCGGCGACCGCGAGCGGCTGCGGCTACATGCGGTCAACCAAGCTCGGAGGACCCGCA 1007849

```

QY 365 TGCAGAGACATCGAGCTTTTGTGATGATGTCATTAACCGGCGCTGTGACGCTC 424
DB 1007848 TCGAAGAGAGCAGCGGTGGTTTCTGTCGATGATGTAAGATCGGTGGCGCATTCAGCTCC 1007789
QY 425 ACCCGGAGGCGCATTTGCGCTTCTGTGTGATTTGCGCAATCCAGATGGGTGGCGCA 484
DB 1007788 ACAACCGAGCGCGCTGCTTCTGTGTGATTTGCGCATCCAGACGAGGTGGAGTTT 1007729
QY 485 TGGCCTCGGGGGTTCATCTGCGCATCTCACTTTTGGGAAACCGGCGCGCATATGTTT 544
DB 1007728 TCGCGTCTGGGAGCTGCTGGGCGCATCTCACTGCGCGAGCGCGCGCTGATCGGCT 1007669
QY 545 TCCCTGCTCTCGCGCTGTGAGATTAAACCATGGGCGCTGCTTCCAGACGGGTGTGACG 604
DB 1007668 TTCTGGGACCAAGGGGTCTATGATGTTGCTTATGGCAACCCCTTCCATCCGCGCTCAAA 1007609
QY 605 AGCGGAGATTTGGTGAATACTGTGTGATTGATGAATTTGTGCACTCCAAATTGC 664
DB 1007608 CCGCGGAGATCTACGGCGCGCATGGGATCATCGACGGCGCTGTTGACATGACCGGCTAC 1007549
QY 665 GTGCAAGCGGTGGCAAAACCTCAAGTTATTCAGCGCGGTAGAGGCAACGATCTTTT 724
DB 1007548 GACCATGCTGATCGTGTGAGGTGCTCATCGACGCTCCGAAACGCTTCCGCGCAC 1007489
QY 725 CTCAACAACCTCTGCGCTGTGCA---CTTCCGATGATGAGGCGAATTGCGGCTTCTGCTG 781
DB 1007488 CGCAACCGCGCGCGCGCGCGTACCAGATGTGCCACAGTGGGACTCGGTGTGATGCGCGC 1007429
QY 782 ACCCGCAGAGGCTTGAAATCGGAGGATTTATGAAACGTTGGGAGGAGACGTGTCAAGC 841
DB 1007428 GGGCGGACCGGCGCGCGCGCTCAGGAGCTACTGCAACGCGCGCACCGAGGTTGT 1007369
QY 842 TTTCTGTGCGGCTGTGCGCATGAGCGCGCTGTGTGCGCTTCCCTGGCGCGCATCG 901
DB 1007368 TGTCAAGAACCGATCAAGGCGAAGCG---GCGACCAACGCTGTGCGCGCTGCGCGCTTTG 1007312
QY 902 GGGGCGCGCGCGCTGTGCTGATTGGGCGAG-----GATCGCGCTTCAAGC 946
DB 1007311 GCGGCGCAACCGCATGTGTCTCGCGCAAGAAAGGCGAGTACGCGCGGAGAGACACTG 1007252
QY 947 TTGGGCGCAGAGCTGCGTTTGTGCGCTGTGTGCAATTTGCTGCGCGCGAGCTAAAC 1006
DB 1007251 TCGGCGCGCGCTGTGCTTACGCGAAGCCGACCGGGATGGCGCTCGCGCGAGCTGTGC 1007192
QY 1007 TCCCATGTGTCTCATTCACACCTCGCGCGCGCAATTGTGCAAGCGGCTGAGAGC 1066
DB 1007191 TGCCTGTGTGTGTGTCAATTGACGCGCGGACCGCGCTGTGCGCGCGAGCGAAGCAG 1007132
QY 1067 TCGGCGTGAAGCTCGAATTGCGCGCACTTGTCAAGCTTATTCAGACGCTCCCTCCCA 1126
DB 1007131 GCGGCGCTGCGCGCAAGTCCGCAATGCTGTGCGGAGTCTCTCACTGTGATACCCCGA 1007072
QY 1127 CCGTTTCCGATATTATTTGATGAGGCGTGTGCGGCGTGTGCGCATGTGCGCGCG 1186
DB 1007071 CCGTGTGATCTGTGTGCGCGCAAGGCGAGCGCGCGCGCTGTGCGCATGTGCGCGCG 1007012
QY 1187 ATCTGTCTAGCGCGCGCAAAACGCGTGTGTGCGCATGTGCGCAAGAGGCGCGCTGCG 1246
DB 1007011 ACCGGGTCTGTGCGCATCTCACTGCTGTGCGCGCGCTGTGCGCAAGAGCGCAAGC 1006952
QY 1247 CCATCTCTTCCGCGCGCAACCAACGCGCGGGAATCATAGAGGAGCAAGCGCTGTGACG 1306
DB 1006951 CGATCTGTCTCGAAGACCTGCTCATGTGCGCGCAACTGCTGTGCGCGCGCGCATCTCGGT 1006892
QY 1307 CGCAGCACTTTTAAAGGCGCTTATCGACGCGATGTGTGCGCGCAACCGAGCACTTTG 1366
DB 1006891 CGGCGGACCTTACTGAAGTCTGGGGAATTGTGACACATGTGTGCGGAGTACCCCGACGCG 1006832
QY 1367 TTGAAGA 1373
DB 1006831 CAGACGA 1006825

```

```

RESULT 8
AA199682/C
ID AA199682 standard; DNA; 4411529 BP.
XX
XX AA199682;
AC 15-JAN-2002 (first entry)
XX
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX Mycobacterium tuberculosis; patient treatment; epidemic monitoring; ds.
XX Mycobacterium tuberculosis.
XX
XX US6294328-B1.
XX
XX 25-SEP-2001.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ.
XX
XX Claim 3; SEQ ID NO 1; 3bp + Sequence Listing; English.
XX
XX
XX The invention relates to evaluating strain variation within and between
XX different populations of the tuberculosis bacterium pathogen,
XX Mycobacterium tuberculosis or related Mycobacterium by determining the
XX nucleotide sequence of the first strain at positions in the complete
XX sequence of the genome that correspond to positions that differ in the
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
XX H37Rv (AA199682). The method is useful for evaluating strain variation of
XX M. tuberculosis and has valuable application in the fields of
XX tuberculosis genetics, epidemiology, patient treatment and epidemic
XX monitoring.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
XX Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;
XX
XX
XX Query Match 21.2%; Score 312.6; DB 22; Length 4411529;
XX Best Local Similarity 54.0%; Pred. No. 5.4e-77;
XX Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
XX
XX
XX 65 ACTCGTTTGGACCTGACAGCTTCAATTTTGGAAATGAATACTCCCAATATGACAAC 124
DB 1008147 ACGGAGTGTGACCGGAGATCTTTCGACGTGAGTATGAGACCTGCGGCTGCGG 1008088
QY 125 TCAATCAAGGTTATGAGAGACCTTTGAGCGGCTCGAAGACGCAATATGATGAT 184
DB 1008087 TAGCGACTCTATGACCGGAGCTGCGCGCTGCGGCGGCAACCGCGCGAGCAAT 1008028
QY 185 CGTAAATTAATCGAAGAGGACCGGTGAGAGGCAATTCGGTATGCGGTTATTTGCGGATT 244
DB 1008027 CGGTGACAGCGGAGAGGAGCGGATTTCCGCGCGGAGGCGGCTGTGTGCTGTGAGT 1007968
QY 245 TTTCTTCTCGGCGGTTTCTTTTGGGACGCTGCGGTGTGCGGATCATGAAGCGGATT 304
DB 1007967 TCGACTTCTGCGGCGGCTCGATTGTGGGTGGGAGGCGGAGATCAACGCGCGCGCTG 1007908

```

```

QY 305 ACCGCGCACAGAGCTGAAACTCCCACTGCTGTCCTCCCTCTCTCCGATGATGCGCGCA 364
Db 1007907 AGCGGGCCACCGCCGACGCGGCTGCCGCTACAGCTCCGAGAGCCACCGCA 1007848
QY 365 TGCAGAAAGCAATCAGAGCTTTTGTCAATGATGTCATTAACCGCGGCTGTGACGCTC 424
Db 1007847 TGCAGAAAGCAAGCTCCGCTTTTGTCAAGATGTCATTAACCGCTCCGCTCCATCCGCTGC 1007788
QY 425 ACCGCGAAGGCGCATTTTGGCGTTTCCGATGATTTGGCAATCCCAAGATGGGCGCA 484
Db 1007787 ACAACGAGGCGCGCTGCCCTACCTGGTATTTTGGCATTCGACACAGGATGAGTTT 1007728
QY 485 TGGCTCTGAGGAGTTCATCTGAGCATCTCACTTTTGGCGAAACCGCGCGCAGATAGTT 544
Db 1007727 TCGGCTCGTGGGCGCTCGCTGGGCGATCTCACGCTGCGGAGCGCGGCGCTGATCGGCT 1007668
QY 545 TCTTGGTCTCTGCGCTGAGTTAACCACTGCGCATGCGCTTCCAGACGCTGTGACG 604
Db 1007667 TTTCTGGGACCAAGGCTCTATGAGTTGCTATGAGCGACCCCTTCCATCCGCGCTCAAA 1007608
QY 605 AGCGGAGAAATTTGTAAGAAATCGGTGATGATGAAATTTGTGCGCACTCCAAATTGC 664
Db 1007607 CCGCGAAGATCTACGCGCGCATGAGATCAGCGGCTGTGCACTGACCGGCTAC 1007548
QY 665 GTGCAAGCGGTGCAAAAAACCTCAAGTTATTCAAGCGGTAGAGGCAACGATCGTTTTT 724
Db 1007547 GACCGATGCTGATGATGCTGTGTTGACGAGTCTATGAGACGCTCCGACCGCTCCGGCAC 1007488
QY 725 CTCCAACAACCTCTGCGCTGCA---CTTCCGCTATGAGAGCGATTTGCGCTTCTGCT 781
Db 1007487 CGCAACACCGCGCGCGCCGACCGCATGTCGCCACATGAGGACCTCGGTGAGCATCGCGCC 1007428
QY 782 ACCCGCAGAGGCTGGAATCGGGGAGATTATGGAACGTTGGGGGCAACGTCGTCAGC 841
Db 1007427 GCGCGACCGCGCGCGCGCTGACAGCTACCTGCAACGCGCGCACCGAGGCTGTGT 1007368
QY 842 TTTCTGTGTCGCGCTGCTGCGCATTTAGCCCGCTGTGCGCTTCCCTGCGCGCATCG 901
Db 1007367 TGTCAAGAAACGATCAAGCGCGAAGG---GCGACCAAGCTGTGCGCGCTGCGCGCTTGG 1007311
QY 902 GGGGCGCGCGCTGCTGATTTGAGC-----GATGCGCGCTTACGCG 946
Db 1007310 GCGGCAACCCAGGTGCTCTCGGCGCAAGGCGAGTAGCGCGCGGGGAGACACTG 1007251
QY 947 TTTGGGCGGAGAGCGTGTGCTTTTGGCGCTGCGCATTTGCTGCGCGCGCAAGTAAAC 1006
Db 1007250 TCGGCGCGCGCTGCTTACGCGAAGCCCGAGCGGAGATGCGCTCGCGCGAGCTGTGC 1007191
QY 1007 TCGCATCGTGTCTCATCATCGACACCTCCGCGCGCAATTGTGCGAGCGCGTGAAGAC 10066
Db 1007190 TGCCTGCTGCTGCTGCTCATTTGACGCGGCGGACCGCGCTTGTGCGCGCGAGCGAAGG 1007131
QY 1067 TCGGATGCGAAGCTGATTTGCGCGACCTTGTCAAGTTATCGACGCTCCCTCCCA 1126
Db 1007130 GCGGCGTGGCGCGCGCAATGCGCATTTGCTGCGCGAGCTCTCTACGCTGATACCCCA 1007071
QY 1127 CGTTTGGTCAATTTGTCAGGGGCTTTGGGGTGGCGCGTGGCGCATNGCTGCCGCGCG 1186
Db 1007070 CCGTGTGCAATCTGCTGGCGCGAGGAGCGGCGCGCGCGCTGGGATGTTGCCCGCG 1007011
QY 1187 ATCTGCTACCGCGCGCAAAACGCGTGGCTTCCGATTTGCCACAGAGCGCGCTCGG 1246
Db 1007010 ACCGGGTGTGCGCGCATCTCCACGCGTGGCTGCGCTTGTCTCCGAGAGACGACAG 1006951
QY 1247 CCATCTCTTCCGCGACACCAACGCGCGGAAATCAATGAGCGACAGGCGTGCAG 1306
Db 1006950 CGATCGTGTCCGAGACACTGCTCATGCGCGCAACTCGCTGCGCGCAAGCATCCGT 1006891
QY 1307 CGCAGCACTTTAAGCCAGGCTTATGACGGATTCGTCGCCAGAAACGAGCACTTTG 1366
Db 1006890 CGGCGACCTACTGAAGTCGGGATTTGTGACACCAATCGTGCAGATACCCCGACGCG 1006831
QY 1367 TTGAAGA 1373

```

```

Db 1006830 CAGACGA 1006824

RESULT 9
AAS54163
ID AAS54163 standard; DNA; 873 BP.
XX
AC AAS54163;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #294.
XX
KW Antisense; ds: prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELITRA) ELITRA PHARM INC.
XX
PI Haasebeck R, Ohlsen KH, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR MPI; 2001-611495/70.
XX
DR P-P8DB; AAU36304.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7800; 511pp; English.
XX
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 873 BP; 152 A; 285 C; 276 G; 160 T; 0 other;

```

```

Query Match 8.3%; Score 122.8; DB 23; Length 873;
Beet Local Similarity 52.8%; Pred. No. 1.5e-24;
Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;
QY 179 ATGAATCGTAATTACTGAGAAAGCAACGTGAGGCGCATTCGGTAGCCGTTATTTTGT 238

```

```

Db      326 AGGACCGCTGATCGCCATGAGCGCAAGCTGACGGGCAATGCGGATGTCGCTTGCCCT 385
QY      239 CCGATTTTCTTCTCTCTGCGGCTTCTTTGGGCACGGTCGGTGGCGCATATGAAG 298
Db      386 TCGATTTCTCTTCTTATGAGGCGGTTGCAATGGCGCATTCGCGGAGCGCTTCTGCGG 445
QY      299 CGATTCACCGCGCCACAGAGCTGAAATCCCACTGCTGCTCTCCCTGCTTCCGGTGTG 358
Db      446 CAGCCAAAGTCGCTTGTAGAGAGCGCTGCGCTGATCTGCTTCTCCGCTCCGGCGG 505
QY      359 CCGCATGAGAGAGCAATTCAGCTTTTGTCTATGATGTCATATACCGCGCTGTGC 418
Db      506 CCGCATGAGAGAGCTGATCTCGCTGATGAGAGCAAGACCTCGCGCTTCTG 565
QY      419 AGCGTACCGCGAGCGCATTTGCGCTGCTGATGATTTGGGCAATCCACAGATGGTG 478
Db      566 CCGCTGCGGCAAGAGCAATCCGTTGCTGCTGATTTGACCGACCGGCTTACGGG 625
QY      479 GCGCCATGCGCTGCGGCTTCACTGCGCATCTCACTTTGGGGAACCGGCGGCGAGA 538
Db      626 GCGTTTCCGCGCAGCTGCGGAGTCTGCGGAGCTGATCGTCCGCAACCCAGGCGCTGA 685
QY      539 TAGGTTCTGAGTCTCGCGTGTGAGTTAACCATGCGGCAATGCGCTTCCAGACGTG 598
Db      686 TCGGCTTCCGCGCTCTCGCTGATGAGACAGACGCTCCGCGAAGCTTCCGGAAGGCT 745
QY      599 TGCAGAGCGGAGAAATTTGTGAAACCTGCTGATTTATGAAATTTGTGTGCCACTTC 658
Db      746 TCCAGGTATGACGATGCTCTTCTTGAACATGCGCCATGACATGATCGTATCGCG 805
QY      659 AATTGCGTGCAGCGGTGCAAA 680
Db      806 AGTTGGGCGCGGCTGCGCAA 827

```

RESULT 10

ABZ39000
ID ABZ39000 standard; DNA; 876 BP.

```

XX      AC      ABZ39000;
XX      DT      07-MAR-2003 (first entry)
XX      DE      N. gonorrhoeae nucleotide sequence SEQ ID 2589.
XX      KW      Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX      OS      Neisseria gonorrhoeae.
XX      PN      WO200279243-A2.
XX      PD      10-OCT-2002.
XX      PE      12-FEB-2002; 2002WO-IB02069.
XX      PR      12-FEB-2001; 2001GB-0003424.
XX      PA      (CHIR-) CHIRON SPA.
XX      PI      Fontana MR, Pizza M, Massignani V, Monaci E;
XX      DR      MPI; 2003-058415/05.
XX      P-PSDB; ABP78030.
XX      PT      New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX      FT      medicament for treating or preventing N. gonorrhoeae infection -
XX      PS      Disclosure; Page 377; 815pp; English.
XX      CC      The present invention relates to proteins from Neisseria gonorrhoeae.
XX      CC      Also disclosed are the nucleic acid molecules encoding the proteins and
XX      CC      antibodies that specifically bind to the proteins. The composition

```

CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC molecules of the invention.

XX SQ Sequence 876 BP; 183 A; 216 C; 272 G; 205 T; 0 other;

Query Match 8.3%; Score 122.2; DB 25; Length 876;
Best Local Similarity 51.6%; Pred. No. 2.2e-24;
Matches 280; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

```

QY      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCCAGAGCAAGCCCAATGCCAT 180
Db      274 AAAGACAGCAAAATATCCGACCGCTTTGATGTCGCGACCGACCTAACCGGGAAGT 333
QY      181 GAATCGTAATTTACTGAGAAAGCACCGTGAAGGCAATTCGGTACCGTTATTTTGTCC 240
Db      334 GACGCGCTGTGTGATGAGAGGAGATGAAACGGTCTGCCGCTGTGTCGTCGCGCTT 393
QY      241 GATTTTCTCTCTCTGCGGCTTCTTTGGCACGCTCGCTGTCGTCGATCATGAGCG 300
Db      394 GAATTCGCGCTTATTCGCGGCTTCGATGCGTTCTGTTGCGGCGAGCGCTTGTGACAGGAT 453
QY      301 ATTACCGCGCACAGAGCTGAAACTCCACGCTGTCTCCCTGCTTCCGGTGTGCG 360
Db      454 GTCGCGCTGTGCTGCGCTGCGATTAATTTGCTGTGTCTGTGCGCGCTTCCGCGCG 513
QY      361 CCGATGACAGAAACAAATGAGCTTTTGTCAATGATGTGTCATTAACCGCGCTGTGAG 420
Db      514 CGTATGACAGAGGGGTAACTCGCTGATGAGATGAGAAACCAAGTCCGCGCTGAT 573
QY      421 CGTACCGCGAGGCGCATTTGCCGTTCTGTGTAATTTGCCCATTCACAGATGGTGGC 480
Db      574 TTGCTGACGAAAAACCGCTCGCTTATTCGCTGTGACCGATCCACTATGAGCGCG 633
QY      481 GCATGCGCTGTGCGGCTTCACTGCGGCACTCACTTTTGGGAAACCGGCGCGAGATA 540
Db      634 GTATCTGCACTTCCGCAATTTTATGCGATTCGCTTGGCAACCGAAGCGCTGATC 693
QY      541 GATTTCTGAGTCTCTGCGTGTGAGTTAACCATGCGGCAATGCGCTTCCAGACGCTGTG 600
Db      694 GATTTTCCGCTGCGCGGCTGATTTGAGACAGCGGTGCGCAAGCGCTCCGGAAGGCTTC 753
QY      601 CAGCAGCGGAGCAATTTGCTGAAATCTGTGTGATGATGGAATTTGTCTCCACTCCAA 660
Db      754 CACGCGCGGAGTTTCTGTGAAAAAGGCGGATGACACAGATTTGACCGCGCGAT 813
QY      661 TTG 663
Db      814 ATG 816

```

RESULT 11

AAS53191
ID AAS53191 standard; DNA; 867 BP.

```

XX      AC      AAS53191;
XX      DT      13-FEB-2002 (first entry)
XX      DE      Enterococcus faecalis DNA for cellular proliferation protein #619.
XX      KW      Antisense; ds; prokaryotic cellular proliferation gene;
XX      KW      antibiotic; antibacterial; drug design.
XX      OS      Enterococcus faecalis.
XX      PN      WO200170955-A2.
XX      PD      27-SEP-2001.
XX      PF      21-MAR-2001; 2001WO-US09180.

```


Db 5321 GATGAAGCTGTCTTAAACGGAGAACAAATTAAGAGACAACTGTGTGCAATTTGGAATC 5380
 QY 238 TCCGATTTTCTTCTCCGCGGTTCTTTGGGACCGGTGCGTGGGATCATGAAAG 297
 Db 5381 ATGAGCGCTAATTTTATCATGAGGAGTATGGAACATTTTGTGTAATAATCAACGC 5440
 QY 298 GCGATTACCGCGCAACAGAGCTGAAATCCCACTGCTGCTCCCTGCTCCGCTGTG 357
 Db 5441 TTGTTGAGCGGCGACAGAAAAGCATTTACAGTAGTAGTTTCACTGATCTGGTGGT 5500
 QY 358 GCGCGATGAGGAAGACATCAAGCTTTGTATGATGTGTCCATTAACCGCGCTGTG 417
 Db 5501 GCCCGATGCAAAAGAAATTTTTCATTTGATCAATAGCGAAAATTTGCGCGCTTGG 5560
 QY 418 CAGCGTACCGGAGGCGCATTTGCGGTTCTGTGATTTTGGGCAATCCAGATGGGT 477
 Db 5561 CAACGGCATTAACAAAGCAGGCTTGTGATCTTAACGATTTGCTGATCCAAAGACTGGC 5620
 QY 478 GCGCGCATGCGCTCGTGGGTTTCATCTGAGCATCTCACTTTTGGGACCGCGCGCAG 537
 Db 5621 GCGTTACCGCAAGTTTGGATGATGGGATATATTTTGGCAAGCCTCAGAGTTTA 5680
 QY 538 ATAGGTTTCTGCTGCTCGGCTCGGCTGTGAGTTAACCACTGGCGATGCGCTTCCAGCGGT 597
 Db 5681 ATCGGTTTGTCTGGCGCGCGTGAATTAAGAAACAAACGATTCGTCAAGAGTTGCCAGATGAT 5740
 QY 598 GTGACAGGCGGAGAAATTTGTGAACATCTGTGTGATGATGAGATTTGTGTGCCACTC 657
 Db 5741 TTTCAAAAGCGCAGATTTCTTTAGAACATGTTTGTGATCAGATTTGTGCCAAGAAAT 5800
 QY 658 CAATTGCGT 666
 Db 5801 CTTTGGCGT 5809
 RESULT 13
 ABS99190
 ID ABS99190 standard; DNA; 6021 BP.
 AC ABS99190;
 DT 18-DEC-2002 (first entry)
 DE Enterococcus faecalis contig sequence #458.
 XX
 XX Computer readable medium; Enterococcus faecalis; microbe; growth;
 KM pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KM therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KM biotech technology; antibacterial; modulator of nucleic acid expression;
 KM contig; ds.
 OS Enterococcus faecalis.
 XX
 XX US2002120116-A1.
 PN 29-AUG-2002.
 PD 29-AUG-2002.
 XX
 XX 04-MAY-1998; 98US-0070927.
 PF 04-MAY-1998; 98US-0070927.
 XX
 XX 04-MAY-1998; 98US-0070927.
 PR
 XX (KUNS/) KUNSCH C A.
 PA (DILL/) DILLON P J.
 PA (BARA/) BARASH S.
 XX
 XX Kunsch CA, Dillon PJ, Barash S;
 PI
 DR WPI; 2002-750065/81.
 XX
 XX Computer readable medium having recorded on it a Enterococcus faecalis
 PT nucleotide sequence useful for detecting diseases related to
 KM Enterococcus infections in animals
 XX

PS Claim 1; Page -; 119pp; English.
 XX
 CC The present invention relates to a new computer readable medium with an
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to
 CC diagnose the presence of E. faecalis in a sample or determining the
 CC presence of a specific microbe in a sample. The invention is also useful
 CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine
 CC to confer resistance to Enterococcal infection, for commercial,
 CC therapeutic and industrial purposes, and for fermenting a particular
 CC sugar source or to produce a particular metabolite. The invention is
 CC useful for detecting diseases related to Enterococcus infections in
 CC animals, and for detecting E. faecalis using biotech technology. The
 CC present nucleic acid sequence represents an Enterococcus faecalis contig
 CC DNA sequence of the invention
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification but was obtained in electronic format directly
 CC from USPTO at <http://seqdata.uspto.gov>.
 XX
 SQ Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;
 Query Match 8.2%; Score 121; DB 24; Length 6021;
 Best Local Similarity 53.0%; Pred. No. 1.1e-23;
 Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
 QY 178 GATGAATGCGTAATTAATGAGAGAGCAACCGGAGGGCATTCGGTACCGCTTATTTG 237
 Db 5321 GATGAAGCTGTCTTAAACGGAGAACAAATTAAGAGACAACTGTGTGCAATTTGGAATC 5380
 QY 238 TCCGATTTTCTTCTCCGCGGTTCTTTGGGACCGGTGCGTGGGATCATGAAAG 297
 Db 5381 ATGAGCGCTAATTTTATCATGAGGAGTATGGAACATTTTGTGTAATAATCAACGC 5440
 QY 298 GCGATTACCGCGCAACAGAGCTGAAATCCCACTGCTGCTCCCTGCTCCGCTGTG 357
 Db 5441 TTGTTGAGCGGCGACAGAAAAGCATTTACAGTAGTAGTTTCACTGATCTGGTGGT 5500
 QY 358 GCGCGATGAGGAAGACATCAAGCTTTGTGATGATGATGATGATGATGATGATGATGAT 417
 Db 5501 GCGCGATGAGGAAGACATCAAGCTTTGTGATGATGATGATGATGATGATGATGATGAT 477
 QY 418 CAGCGTACCGGAGGCGCATTTGCGGTTCTGTGATTTTGGGCAATCCAGATGGGT 477
 Db 5561 CAACGGCATTAACAAAGCAGGCTTGTGATCTTAACGATTTGCTGATCCAAAGACTGGC 5620
 QY 478 GCGCGCATGCGCTCGTGGGTTTCATCTGAGCATCTCACTTTTGGGACCGCGCGCAG 537
 Db 5621 GCGTTACCGCAAGTTTGGATGATGGGATATATTTTGGCAAGCCTCAGAGTTTA 5680
 QY 538 ATAGGTTTCTGCTGCTCGGCTCGGCTGTGAGTTAACCACTGGCGATGCGCTTCCAGCGGT 597
 Db 5681 ATCGGTTTGTCTGGCGCGCGTGAATTAAGAAACAAACGATTCGTCAAGAGTTGCCAGATGAT 5740
 QY 598 GTGACAGGCGGAGAAATTTGTGAACATCTGTGTGATGATGAGATTTGTGTGCCACTC 657
 Db 5741 TTTCAAAAGCGCAGATTTCTTTAGAACATGTTTGTGATCAGATTTGTGCCAAGAAAT 5800
 QY 658 CAATTGCGT 666
 Db 5801 CTTTGGCGT 5809
 RESULT 14
 ABX65750
 ID ABX65750 standard; DNA; 676 BP.
 AC ABX65750;
 DT 07-MAY-2003 (first entry)
 DE Helicobacter pylori selected interacting domain (SID) DNA #349.
 XX Protein-protein interaction; ulcer; selected interacting domain;
 KM SID; gene; ds.
 XX

```

XX Helicobacter pylori.
OS
XX WO200266501-A2.
XX
XX 29-AUG-2002.
XX
XX 28-DEC-2001; 2001WO-EP15428.
XX
XX 02-JAN-2001; 2001US-259302P.
XX
XX (HYBR-) HYBRIGENICS.
XX (INSP) INST PASTEUR.
XX
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX WPI; 2002-674910/72.
XX DR P-PSDB; ABUS1006.
XX
XX New complexes of protein-protein interactions in Helicobacter pylori,
XX useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals -
XX
XX Claim 7; Page 172; 642pp; English.
XX
XX The invention describes a complex of protein-protein interactions in
XX Helicobacter pylori selected from 421 complexes given in the
XX specification. The complex of protein-protein interactions are useful
XX for screening for agents which modulate the interaction of proteins.
XX Modulating compounds which binds to a targeted bacterial protein may be
XX used for treating or preventing ulcers in a human or animal. This
XX sequence encodes a selected interacting domain (SID), identified via
XX protein-protein interactions.
XX
XX Sequence 676 BP; 183 A; 119 C; 190 G; 184 T; 0 other;
XX
XX
XX Query Match      8.1%; Score 119.2; DB 24; Length 676;
XX Best Local Similarity 50.9%; Pred. No. 1.4e-23;
XX Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
XX
XX
XX 93 TTTCTTGAAATGAATCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGA 152
XX   |||||
XX 73 TTTACGGCCTATGATCTTTAAATTTGGTGATTAAGAGACTATTAACAACGATTA 132
XX
XX 153 GGGGGCTGAAGACAGCCAAATGCGATGATCCGGTAATTAATGAGAAAGACCGTGA 212
XX
XX 133 AAAATAGCAAAAAGAGACTTAACCCCAAGCTCAGTATCAGCGGTGAGGCTAAATCA 192
XX
XX 213 GGGCATTCGGGTAGCCGTTATTTGTCGATTTTCTTCTCGCGCGGTTCTTTGGGAC 272
XX   |||||
XX 193 CCGCATGCTTTGACAGATGCTGTGTTGATTTTACCTTTATGGGGGAGTTAGGCTC 252
XX
XX 273 GGTGCGGTGCGATGATGATGAGGCGATTCACCGCGCACAGCTGAAATCTCCACT 332
XX
XX 253 TGTGAGAGGCGAAAAGATGTAAGAGCAATCATGCGCGGTGCTTAAAGAGAACGCT 312
XX
XX 333 GGTGCTTCCCTGCTTCCGGTGTGCGGCGATGAGAGAAAGCAATCCAGCTTTGTAT 392
XX
XX 313 ATTGATTTGTTAGCGAGTGGGGGGGCTAGATGAGAAAGATCCACTTATTTCTCATGA 372
XX
XX 393 GATGATGTCATTAACCGGGCTGTGCGAGCGTCACCGCGAGCGCATTTGCCGTTCTGAT 452
XX
XX 373 AATGCTTAAACGAGAGCGGGCTTTGACCGATTTGAGTGAAGGCGAAATCTCTTATTTTC 432
XX
XX 453 GTATTGTGCGCAATCCACAGATGGTGGCGCATGCGCTTGCGGGGTTCAATCGGCGATCT 512
XX
XX 433 GGTCTTAAGCGATCCACTTATGAGGGGCGTATGCGCATCTTTGCTTTTATGAGGATCT 492
XX
XX 513 CACTTTTGGGGAACCGCGCGCGCAATAGTTCCTCGGCTCTCGCGTGTGAGATTAAC 572
XX
XX 493 CATTATCGAGAGCCAGAGGGCGATGATGAGCTTTTGGGGGCGCTAGGGTATTAACCAAC 552
XX
XX 573 CACTGGGAGATGCGCTTCCAGACGATGTGACAGAGGCGAGAAATTTGGTAAATCTGTGT 632
XX

```

```

DB 553 TATAGGGCGGATTTGCTGAGGCTTCAACAGCGGAATTTTATTAAGCATGCTT 612
XX
XX 633 GATTGATGAATTTGTG 648
XX
XX 613 GATTGATATGATTTGTG 628
XX
XX
XX RESULT 15
XX AAS33651
XX ID AAS33651 standard; DNA; 870 BP.
XX
XX AAS33651;
XX
XX 13-FEB-2002 (first entry)
XX
XX Helicobacter pylori DNA for cellular proliferation protein #105.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX Helicobacter pylori.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELITRA) PHARM INC.
XX
XX Haeelbeck R, Ohlsen KU, Zykkind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX DR P-PSDB; AAU35792.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 7288; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 870 BP; 253 A; 147 C; 229 G; 241 T; 0 other;
XX
XX
XX Query Match      8.1%; Score 119.2; DB 23; Length 870;
XX

```

Best Local Similarity 50.9%; Pred. No. 1.5e-23;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```

QY 93 TTCTTGGAAATGAATCCCAATATATGCAACCTCAATCAAGCTATTCAGAGACCTTGG 152
   |||||
Db 240 TTACGACCTTAATGATCTTTAAATTTCTGTGATTAAGAGACTATTAACAACGATTA 299
   |||||
QY 153 GCGGCTCGAAGCAAGGCCAATATCGATGATCGTAAATTAATGAGAAAGCACCGTGA 212
   |||||
Db 300 AAAATAGAAAAAAGAGCTAACCGCCCACTAGTATCAGCGGTGAGGCTAAATCA 359
   |||||
QY 213 GGGCAATCCGGTACCGGTATTTTGTCCGATTTTTCCTTCCTCGCGGTTCTTTGGGAC 272
   |||||
Db 360 CCGCATGCCCTTTCAGATCGTGTGTTGATTTTACCTTATGAGGGGAGTTAGGCTC 419
   |||||
QY 273 GGTGCGGTGGGTGCGGATCATGAAGCGATTCACCGGCCACAGAGCTGAATCCCACT 332
   |||||
Db 420 TGTGAGAGGCGAAAAAGATGTAAGACCAATCAATCGCGGTCTAAAAGAGAGCGTT 479
   |||||
QY 333 GCTGTCTCCCTCTTCCGCTGTGCGGATGCGAGAGAGCAATCGAGCTTTGTCA 392
   |||||
Db 480 ATTGATTTGTTTGAAGAGTGGGGGGCTAGGATGCAAGATCACTTATTCGCTCATGCA 539
   |||||
QY 393 GATGTGTCCATTAACCGCGGCTGTGCAAGCTCACCGAGGCGCATTTGCCGTTCCGTG 452
   |||||
Db 540 AATGCTTAAACGAGGCGGCTTTGAACGATGAGTGAAGGCCAATCCCTTTCAATTTC 599
   |||||
QY 453 GTATTTCGCCAATCCACGATGAGTGGCGCATGCGCTCGTGGGGTTCAATCTGGCATCT 512
   |||||
Db 600 GCTCTTAAGCGATCCCACTTAATGGGGCGCTTAAGCGATCTTTTGTGCTTTTAAAGGGATCT 659
   |||||
QY 513 CACTTTTGGCGAACCAGCGCGCAGATAGGTTTCTGAGTCTCGCGTGTGAGTTAAC 572
   |||||
Db 660 CATTAATCGAGAGCCAGGGGCGATGATAGGCTTTGGCGGGCTTAGGGTATTAAGCAAC 719
   |||||
QY 573 CACTGGGCATGCGCTTCCAGACGCTGTGACAGCGCGAGAAATTTGGTGAACCTGTGT 632
   |||||
Db 720 TATAGGGGCGGATTTGCTGAGGGCTTTCAACAGCGGAATTTTATAGAGCATGCTT 779
   |||||
QY 633 GATTGATGGAATTGTG 648
   |||||
Db 780 GATTGATGATTTGTG 795

```

Search completed: November 12, 2003, 20:26:11
Job time : 440.983 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 18:57:58 ; Search time 3242.57 Seconds
(without alignments)
11040.790 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtgagaagcgtttccgac.....gatttcagcgtttcagcag 1473

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_num:*
18: em_ges_inv:*
19: em_ges_pin:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrl:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	6.0	488	28	BH379574 AG-ND-133
2	86.2	5.9	707	13	BH654446 1112113F0
3	84.2	5.7	604	9	AMS84783 N210917e
4	82.6	5.6	773	12	B1311746 EST531349

5	81.2	5.5	1063	29	B2548722 pac61-60
6	81	5.5	712	12	B1310727 EST531247
7	79.8	5.4	580	14	CA990607 EST644115
8	79.8	5.4	792	10	BC584921 EST486683
9	78.8	5.3	1042	29	B2561247 pac62-164
10	78.4	5.3	539	12	B1725661 1031080C0
11	77.6	5.1	652	29	AG247247 Lotus Jap
12	75	5.1	617	14	CB977199 CAB40003
13	74	5.0	668	9	A1229631 EST226326
14	73.8	5.0	677	28	B2035752 oeh49c05
15	73.6	5.0	606	14	CB346970 CAB2SG000
16	73.4	5.0	453	14	CB081454 hks1h06.9
17	73.4	5.0	708	28	B2051213 jnr62h07
18	72.2	4.9	675	28	BH988285 oej57b0c
19	72.2	4.9	696	28	B2012471 oed33c11
20	72.2	4.9	700	28	B2033778 oei10d09
21	72.2	4.9	703	28	B2017314 oei67c12
22	72.2	4.9	704	28	BH930860 odj01b11
23	72.2	4.9	709	28	B2075724 1k153c01
24	72.2	4.9	711	28	AO957569 LERA046TF
25	72.2	4.9	752	28	B2033142 oeh45a09
26	72.2	4.9	781	28	B2071328 1jul3h07
27	72.2	4.9	807	28	BH715901 BOHYJ64TR
28	72.2	4.9	844	28	BH462655 BOHJR54TF
29	71.8	4.9	1872	28	BH770899 LAMCtag63
30	71.4	4.8	607	14	CB977121 CAB40003
31	71.4	4.8	637	28	AO964265 LERQ88TF
32	70.4	4.8	843	28	BH477984 BOHQP36TF
33	70	4.8	719	28	B2052334 jnr68b12
34	69.4	4.7	697	28	B2006948 oem84f09
35	69	4.7	405	28	BH397912 AG-ND-143
36	68.8	4.7	627	9	AU293522 AU293522
37	68.4	4.6	692	28	B2026810 oeh37c06
38	68.4	4.6	748	28	BH987475 oej17e06
39	68.2	4.6	523	28	AO991181 RfC02065
40	68	4.6	698	28	BH994577 oei62f03
41	67.8	4.6	452	9	AMS87803 ST67B11 P
42	67.4	4.6	722	28	BH925341 oei46f02
43	67	4.5	886	29	B2553868 pac61-60
44	66	4.5	1333	29	B2579621 ms82_668
45	65	4.4	558	12	B1725648 1031080B1

ALIGNMENTS

RESULT 1
BH379574
LOCUS AG-ND-133C23, 488 bp DNA linear GSS 10-DEC-2001
DEFINITION 'genomic survey sequence.'
ACCESSION BH379574
VERSION BH379574.1 GI:17325716
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 488)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren
, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J.
and Collins, F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
JOURNAL MEDLINE
PUBMED 12655398
COMMENT Other_GSSs: AG-ND-133C23.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers

FEATURES
source

1..488
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-13C23"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 137 a 86 c 109 g 156 t
ORIGIN

Query Match 6.0%; Score 88; DB 28; Length 488;
Best Local Similarity 49.0%; Pred. No. 7.2e-13;
Matches 235; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

132 AGGCTATGAGAGACCTTGGAGCGGCTCGAGCAAGGCAATGCGATCGGTAT 191
1 AGCTTATCTGACCCATTAAGAGGTAAAGGCAAGACTAATTAACGATTCGTACG 60
192 TACTGAGAAGGACCCGTGAGAGGCAATCCGCTAGCCGTTATTTTTCGATTTTCTT 251
61 TAACTCTGTGTGTAAGTAAGCAAGCACTGGAATTTGTTTATGATGATTTCTCTT 120
252 CCTCGCGGCTTTTGGGACCGTGGCGGTGGTGGCATGAAAGCGATTCACCGGC 311
121 TATCGAGAGCTCATATGCTCTGTATGAGAAATAATCATCCGCTATCGACTATGC 180
312 CACAGAGTGAACCTCCAGCTGCTGCTCCCTGCTCCGCTGGTGGCGCATGACAGA 371
181 AATGAGATTAATTAACATTAATTAATTTTCCAGTCTGAGCAAGATGACAGA 240
372 AGACATGAGCTTTTGCATGATGATGTCATTAACCGGCGCTGTCAGCGTCAACGCGA 431
241 AGACATTAATTAATTAAGCAAGTGGCAAAAGTCAAGCTAAGTCAAGCTTTGCGA 300
432 GCGCATTTGCCGTTCTGCTGATTTTGGGCAATCCACAGATGGTGGCGCCATGCTC 491
301 CAAGGCTTACCATTAATTTGCTTAATTAACCTTAATTTGCTGATTAACAGCTTC 360
492 GTGGGCTTCAATGGGCAATCACTTTTGGGCAACCGCGCGCGAGATAGTTTCTGG 551
361 TTTTCGATGACTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
552 TCTCGCGGTGAGATTAAACCACTGGGAGTCCGCTTCAGACGGTGGTCAGACAGCGGA 611
421 TCTTGAAGATTAACAAGAAATTTGTAAGATCTTCTCGAGGCTTCAGACATCTGA 480

RESULT 2
BU654446 707 bp mRNA linear EST 30-SEP-2002
LOCUS BU654446
DEFINITION 111213P06.y1 C. reinhardtii CC-1690 (mt-), CC-1691 (mt-), Gamete (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA
ACCESSION BU654446
VERSION BU654446.1 GI:23366627
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.

REFERENCE 1 (bases 1 to 707)
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre

TITLE P., McDermott, J.P., Shlager, J., Siffow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1112
Unpublished

JOURNAL Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source

1..707
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="21gr (CC-1690 wild type mt+) & 6145C (CC-1691 wild type mt-)"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"
/note="Vector: pLuscript II SK-; Site_1: EcoRI; Site_2: XhoI; Gamete library was constructed by Hui Zhao, Min Lu, Jeffrey McDermott, William J. Snell and John Davies. Strain 21gr cells (CC-1690; mating type plus) and strain 6145C cells (CC-1691; mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Sager and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period. PolyA mRNA was purified from each sample every 2 hours for the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pLuscript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 127 a 241 c 227 g 112 t
ORIGIN

Query Match 5.9%; Score 86.2; DB 13; Length 707;
Best Local Similarity 50.6%; Pred. No. 2.5e-12;
Matches 208; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

942 CACGCTTGGCGCGAGAGCTGCTTTGGCGCTGTCGATTTGCTGGCGCGAGCT 1001
286 CATGCCCGACCGCAACGCTACCCCAAGGCGCTGCTTACATGCGCACCGCAAGTT 345
1002 AAACCTGCGCATTCGTGCATCATGACACCTCCGCGCGGAATTGTGCGAGCGGCTGA 1061
346 TGTCTGCGCATTCATCCTTCTGTGAGACGCCCGAGACCTATCCGCGAATCCGGGA 405
1062 GAGCTCGGCATTCGCAAGCTGATTCGCGCGACCTTGTCCAAAGTTATGACGCTCCCT 1121
406 GGAAGCTGGCGAGGCGAGGCGATTCGCGAATCTCGTGAAGATGTTCCGCTCGCGT 465
1122 CCCACGCTTTCGCTATTAATGTCAGGCGCTGGCGGTCGCGCTGCGCATGCTGCC 1181
466 GCCCATATCTCGGTGTGCTATTCGAGGCGGCTCGCGGCGCGCTGCGCATTTGCTG 525
1182 CGCCGATCTGTCTACCGCGCGCAAAACGCTGCTCTCGCATTTGCAACAGAGGGGCG 1241
526 CGCCACCGCAACTGTATCATGAGAACGCGGTCTACTAGTGGCTCGCGCGGAGGCTG 585
1242 CTGCGCATCTCTTTCGCGACACCAACGCGCGGAATCATAGAGCAACAAGCGGT 1301

Db 586 CGCCGCATCTGTGGAAGCCGCTCTGCGCCGCGGAGGCCACTAGAGCCCTGCGCAT 645
 QY 1302 GCAGGCGACGACCTTTTAAGCCAGGGCTTATCGAGGATCGTCGCCGA 1352
 Db 646 CACCTCGGCGGAGCTGTGTAAGTTGGCCTCATGACACATGTCGCCGA 696

RESULT 3
 AM584783
 LOCUS
 DEFINITION N210917e MHAM Medicago truncatula/Gloms versiforme mixed EST
 library cDNA clone MHAM-7110, mRNA sequence.
 ACCESSION AM584783
 VERSION AM584783.1 GI:7261837
 KEYWORDS EST.
 SOURCE Medicago truncatula/Gloms versiforme mixed EST library
 ORGANISM Medicago truncatula/Gloms versiforme mixed EST library
 Eukaryota; mixed EST libraries.
 REFERENCE 1 (bases 1 to 604)
 AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.R. and Fraser,C.M.
 TITLE ESTs from roots of Medicago truncatula after colonization with
 Gloms versiforme
 JOURNAL Unpublished
 COMMENT Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-223-7380
 Email: mjharrison@noble.org
 Other name: MHAM-7b-E05; Date: 3/14/00; Updated to the database of
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is
 available at "http://chryslie.tamu.edu/medicago".
 Seq primer: 73.

FEATURES
 source
 1..604
 Location/Qualifiers
 /organism="Medicago truncatula/Gloms versiforme mixed EST
 library"
 /mol_type="mRNA"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="MHAM-7110"
 /tissue_type="roots colonized with Gloms versiforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Gloms versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XL0LR"
 /clone_1ib="MHAM"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Gloms versiforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assisit
 helper phage and propagated in XL0LR cells."

BASE COUNT 179 a 95 c 136 g 194 t
 ORIGIN

Query Match 5.7%; Score 84.2; DB 9; Length 604;
 Best Local Similarity 48.8%; Pred.No. 8.3e-12;
 Matches 258; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

QY 127 AATCAAGCTATGACAGACTTGGAGCGGGCTCGAAGCAAGCCAAATGCATGAATCG 186
 Db 13 AATGATGATTAACAAGATGCTTGAATTTTCAAGACAGAACCGATTAAGATGCG 72
 QY 187 GTAATTACTGGAAGACCGCTGAGAGGGCATTCGGGTAGCCGTAATTTTGTCCATTTT 246
 Db 73 GTTCAACAGGACGACGCTCAAGTAATGATTAATCTGTAGCAATAGTATGATTTT 132

QY 247 TCCTTCCTCGGCGTTCTTTGGGACGCGCTCGCTGCTGCGCATGATGAAGCGATTCAC 306
 Db 133 GAGTTTATGGAGAGTATGAGATCGGATCCGTATGGGTGAGAAATACATCTCGTTATGTA 192
 QY 307 CGCGCACAGAGCTGAATCTCCACTGCTGTCTCCCTGCTCCGTTGCGGCGCATG 366
 Db 193 TATGCTACCAATCAACGTTTACTCTTATTAATGATGCGTGTGAGAGCGCGTATG 252
 QY 367 CAGAGACAAATCGACGTTTGTATGATGATGATGCCATTAACCGCGGCTGACGCTAC 426
 Db 253 CAGAGAGAGATTTGAGCTTATGCAAAATGCTAAATTTCTCTTTATATATATAT 312
 QY 427 CGCGAGCGCATTTGCCCTCTCGTGTATTTGCGCA--ATCCACGATGGGTGCGCC 483
 Db 313 CAATCATCAAAAGTATTTCTATCTATACCGATCTTACATCTCCATCACTGATGGGTA 372
 QY 484 ATGGCTTGTGGGGTTCAATCTGGGATCTCACTTTGGGGAACCGGCGCGCATAGT 543
 Db 373 ACAGCTAGTTTGGATGTGGGGGATATCAATTCCTGCGGAACCGATGCTTACATGCA 432
 QY 544 TTCTGGGTCTCGCGTGTGAGTATCACTGCGCATGGCGCTTCAGACGCTGAG 603
 Db 433 TTGCAAGTAAAGATTAATGAAGAAAGTTGAAGATTTGAAGTCCGAAAGTATACAA 492
 QY 604 CAGCGGAGAAATTTGTAAGAACTGCTGATGATGATGATTTGTCCG 652
 Db 493 TCGGTGAATTTTATTCGAAAAGGTCATTTGATTCATCTGATCCAC 541

RESULT 4
 B1311746
 LOCUS
 DEFINITION EST5313496 GRSD Medicago truncatula cDNA clone pGRSD15E20 5' end,
 mRNA sequence.
 ACCESSION B1311746
 VERSION B1311746.1 GI:14986073
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 773)
 REFERENCE Grusak,M.A., Sanac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
 AUTHORS J., U. and Fraser,C.M.
 TITLE ESTs from developing reproductive tissues of Medicago truncatula
 JOURNAL Unpublished
 COMMENT Contact: Michael A. Grusak
 USDA/ARS Children's Nutrition Research Center
 Baylor College of Medicine
 1100 Bates Street, Houston, TX 77030-2600, USA
 Tel: 713-798-7044
 Fax: 713-798-7078
 Email: mgrusak@bcm.tmc.edu
 B1398601e
 TIGR sequence name: MTPA234TK
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA GCG GAT CC).
 Location/Qualifiers
 1..773
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="pGRSD15E20"
 /tissue_type="immature seeds"
 /dev_stage="Immature seeds, 11 to 19 days after
 pollination"
 /clone_1ib="GRSD"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Immature seeds, collected from pods ranging in age
 from 11 to 19 days after pollination, were harvested from


```

Oy 301 ATTACCGCGGCACAGAGCTGAACTCCACGCTGCTGCCCTTCCTCCGCTGTCGCG 360
Db 122 ATGTATATATGTACCAATCAACGTTTACCTTTATATATATATATATATATATATAT 181
Oy 361 CGCATGACAGAAAGAACATCGAGCTTTTGTATGATGTGTCCATCAACCGCGCTGTGAC 420
Db 182 CGTATGACAGAAAGAAAGTTTATGCTTAATGCAAAATGCTAAATTTCTGCTTTATAT 241
Oy 421 CGTACCGCGAGCGGCATTTGCGCTTCTGCTGTATTTGCGCA--ATCCACGATGGGT 477
Db 242 AATTATCAATCAATCAAAAGTTATTTATGATGACGATATCACTTCTTACTACTGCT 301
Oy 478 GCGCCATGCGCTCTGCGGCTTCATCTGCGATCTCACTTTTGGGACCGCGCGCGAG 537
Db 302 GGGGTAAACAGCTAGTTTGGAAATGTTGGGGATATCATTTCTTGGAAACCGCATGCTTAC 361
Oy 538 ATAGATTCTCTGCTGCTCGCTGCTGAGTTAACCTGAGCATGCGCTTCACAGCGGT 597
Db 362 ATAGCATTTGACAGTAAAGAGTATTTGAGAAAGCTTGAAGATGGAAGTCCGCAAGGT 421
Oy 598 GTGACGACGCGGAGAAATTTGGTGAACCTGCTGTGATTGATGAATTTGTGCG 652
Db 422 ATACATCGCGCTGAATTTTATTCGAAAAGGCGTCAATTTGATTCAGTCTGACAC 476

RESULT 8
BGS84921 792 bp mRNA linear EST 11-APR-2001
LOCUS EST84921 MHAM Medicago truncatula/Glomus versiforme mixed EST
DEFINITION library cDNA clone PMHAM-20E21 5' end, mRNA sequence.
ACCESSION BGS84921
VERSION BGS84921.1 GI:13599985
SOURCE Medicago truncatula/Glomus versiforme mixed EST library
MEDICAGO truncatula/Glomus versiforme mixed EST library
Eukaryota: mixed EST libraries.
1 (bases 1 to 792)
Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Uterback,T., Cho,J.
and Fraser,C.M.
ESTs from roots of Medicago truncatula after colonization with
Glomus versiforme, 2001
Unpublished
JOURNAL Contact: Harrison M.J.
COMMENT Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-223-7380
Email: mjharrison@noble.org
Noble EST name: N380393e TIGR sequence name: MTDBW35TK More
information is available at: http://www.medicago.org
Seq primer: SKmod (CTA GAA CTA gtg gat CC).
FEATURES
source 1..792
/organism="Medicago truncatula/Glomus versiforme mixed EST
library"
/mol_type="mRNA"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="PMHAM-20E21"
/life_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0LR"
/clone_lib="MHAM"
/notes="Vector: pBluescript SK-, Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Unizap XR vector from
Stratagene and packaged using GigaPack III Gold packaging

```

```

BASE COUNT 264 a 112 c 151 g 265 t
ORIGIN
Query Match 5.4%; Score 79.8; DB 10; Length 792;
Best Local Similarity 49.5%; Pred. No. 1.4e-10;
Matches 235; Conservative 0; Mismatches 237; Indels 3; Gaps 1;
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

Oy 181 GAATCGGTAATTAATCTGAGAAAGCACCCTGAGAGGCAATTCGGTAGCCGTTATTTGTC 240
Db 33 GATCGCGTTCAACAGGACACAGGTCAGTAAATGATATTCCTGATGCAATAGATATATG 92
Oy 241 GATTTTCTTCTCTCGCGGCTTTCTTTGGGACAGGTGCGCTGCGTATCATGAAGCG 300
Db 93 GATTTTGAATTTAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 152
Oy 301 ATTACCGCGGCACAGAGCTGAAACTCCACTGCTGCTGCCCTGCTCCGCTGTCGCG 360
Db 153 ATGTATATATGTACCAATCAACGTTTACCTTATTTATATATATATATATATATATAT 212
Oy 361 CGCATGACAGAAAGAACATCGAGCTTTTGTATGATGTGTCCATPAACCGCGCTGTGAC 420
Db 213 CGTATGACAGAAAGAAAGTTTATGAGCTTATGCAAAATGCTAAATTTCTGCTTTATAT 272
Oy 421 CGTACCGCGAGCGCATTTGCGCTTCTGCTGTATTTGCGCA--ATCCACGATGGGT 477
Db 273 AATTATCAATCAATCAAAAGTTATTTATGATGATGATGATGATGATGATGATGATGAT 332
Oy 478 GCGCCATGCGCTCTGCGGCTTCATCTGCGCATCTCACTTTTGGGAAACCGCGCGCAG 537
Db 333 GGGGTAAACAGCTAATTTTGGAAATGTTGGGGATATCATTTCTTCCGCAACCGATGCTTAC 392
Oy 538 ATAGATTCTCTGCTGCTCGCTGCTGAGTTAACCTGAGCATGCGCTTCACAGCGGT 597
Db 393 ATACATTTTGCAGGTAAGAGTATTTGAGAAAGCTTGAAGATGCAAGTCCGCAAGGT 452
Oy 598 GTGACGACGCGGAGAAATTTGGTGAACCTGCTGTGATTGATGAATTTGTGCG 652
Db 453 ATACATCGCGCTGAATTTTATTCGAAAAGGCGTCAATTTGATTCACGCTGACAC 507

RESULT 9
BZ561247 1042 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3167.y3 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3167, genomic survey sequence.
ACCESSION BZ561247
VERSION BZ561247.1 GI:27180878
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1042)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..1042
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

```

```

/db xref="taxon:2164"
/clone lib="pac2-164.3167"
/clone lib="pac2-164"
/notes="clinical isolate 2-164 whole genomic shotgun library."
BASE COUNT      207 a      320 c      270 g      244 t      1 others
ORIGIN

Query Match
Best Local Similarity 50.3%; Pred. No. 2.9e-10; Length 1042;
Matches 194; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 970 GCGGCTGTGGGATTTGCTGGCGCGGAGCTAAACCTCCGATCGTGTTCATCATGAC 1029
DB 110 GCGTGGCGGCTATGTAATGCGCGAAGCCTTCAAGATCCCATCTTCACTTCATGAC 169
QY 1030 ACCCTCGGCGCGGATTTGTCGACGCGGCTGAGAGCTTGGCATTCGCAAGCTGATTCG 1089
DB 170 ACGCCCGGCGCTTACCCCGGCGCATTCGATGCCGAGAAACGCGGCGAGAGGCGATCGCC 229
QY 1090 GCGACCTGTTCAGAGCTTATGACGCTCCCTCCACCGCTTTCGTATTTATGTCAG 1149
DB 230 TGGAACTGCGGAGTGTGCGCGGAGTGAAGACGCGCATCATGCGCACCGTATCGCGAG 289
QY 1150 GCGGTTGGCGGTGCGCGCGCTGCGCATGTGCGCGCATCTGGTCTAGCGCGCGGAAAC 1209
DB 290 GCGGCTTCCGCGCGCGCGCTGCGCATGTGCGGTGTGCGACATGTAACATGCTGCAATAC 349
QY 1210 GCGTGGCTGTGCGCATTTGCGACGAGGCGCGCTTGGCGCATCTTCTTCGCGACACCAAC 1269
DB 350 TCCACCTATTCGAGTGTGCTGCGCGGAGGCTGCGCTTCATCTTGAAGAACCGCGAG 409
QY 1270 CACGCGCGGGAATCATATAGGACAGGCGGTGCGAGCGCGCATTTTAAGCCAAAGG 1329
DB 410 AAGGCGCGGAAAGCGCGCGGCGCATGCGGCGCATCACCGCGAGCGCGTGAAGAGCTGGCG 469
QY 1330 CTTATCGACGGGATCGTCGCGGAAAC 1355
DB 470 ATGCTGACAAAGTCAATGACGAGAC 495

RESULT 10
LOCUS      B1725661      539 bp      mRNA      linear      EST 19-SEP-2001
DEFINITION 1031080C09.y1 C. reinhardtii CC-1690, Stress II (normalized),
ACCESSION  B1725661
VERSION    B1725661.1 GI:15701356
KEYWORDS  EST.
SOURCE     Chlamydomonas reinhardtii
ORGANISM  Chlamydomonas reinhardtii; Chlorophyta; Chlorophyceae; Volvocales;
           Eukaryota; Viridiplantae; Chlorophyta; Chlamydomonadales;
           Chlamydomonadales; Chlamydomonadales.
           1 (bases 1 to 539)
REFERENCE  Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
AUTHORS    P., McDermott, J. P., Shrager, J., Sillflow, C. and Stern, D.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
           Unicellular System for Analyzing Gene Function and Regulation in
           Vascular Plants. Project: 1031
JOURNAL    Unpublished
COMMENT    Contact: Charles Hauser
           DCMB Box 91000
           Durham, NC 27708-1000
           Tel: 919 613 8159
           Fax: 919 613 8177
           Email: chauser@duke.edu.

FEATURES
SOURCE     1..539
           location/Qualifiers
           /organism="Chlamydomonas reinhardtii"
           /mol_type="mRNA"
           /strain="CC-1690 wild type mt+ 21gr"

```

```

/db xref="taxon:3055"
/clone lib="C. reinhardtii CC-1690, Stress II (normalized)
Lambda Zap II"
/notes="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with Exsist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."
BASE COUNT      95 a      185 c      175 g      84 t
ORIGIN

Query Match
Best Local Similarity 51.0%; Pred. No. 3e-10; Length 539;
Matches 210; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

QY 942 CACGCTGGGCGCGAGAGCTGCTTGGCGGTGTGCGCATTTGCTGGCGCGAGCT 1001
DB 56 CATGCCCAAGCCCAACGCGTACCGGAGCGGTGCGCTTATGCGCCAGCGCAAGATT 115
QY 1002 AAACCTGCGATCGTGTGCATCATGACACCTCCGCGCGCAATGTTGCGAGCGGCTGA 1061
DB 116 TGTCTGCCATCATCATCTTGTGTGACACGCCCGGAGCTTATGCCGCGAAGCCGCGA 175
QY 1062 GAGGCTGCGCATTCGACAGCTGATTCGCGGACCTTGTCAAGCTTATGACGCTCCCT 1121
DB 176 GAGGCTGGGCGAGGCGGAGCGCATTTGCCGTGAACCTGGTAGATGTTCCGCTCGGT 235
QY 1122 CCCCACCGTTTGGCATTTATTTGTCAGGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1181
DB 236 GCCCATCATCTCGGTGTCATTTGCGAGGCGGCGGTGCGGTGCGGTGCGGTGCGGTG 295
QY 1182 GCGCATCTGTCTTACGCGGCGGAAACCGGTGCTGCCGATTCGACACGAGAGGCGC 1241
DB 296 CGCCAAACCGCAACCTGATCATGAGACGCGGTCTACTACGTGGCTGCCGAGGCGCTG 355
QY 1242 CTCGGCATCTCTTCCGCGA-CACCAACCAACGCGCGGAAATCATAGAGCGACAAAGCG 1300
DB 356 CGCGGCATCTCTGAGGAAAGCGGCTGCGCGCGGCGGAGGCGACATGAGGCCCTGCGCA 415
QY 1301 TGCAGGCGCAGCGACATTTAAGCCAAAGGCTTATGACGCGGATCGTCCGCA 1352
DB 416 TCACCTCGCGGAGGTGTAAGTTCCGCGTCAATGACACATGCGTCCGCA 467

RESULT 11
LOCUS      AG247247      652 bp      DNA      linear      GSS 13-DEC-2002
DEFINITION Locus Japonicus DNA, clone: LjT20K03_not, genomic survey sequence.
ACCESSION  AG247247
VERSION    AG247247.1 GI:26647012
KEYWORDS  GSS.
SOURCE     Locus japonicus
ORGANISM  Locus japonicus
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
           Locus.
REFERENCE  1
AUTHORS    Sato, S., Nakamura, Y. and Tabata, S.
TITLE      Locus japonicus TAC End sequences
JOURNAL    Published Only in Database (2002)
REFERENCE  2 (bases 1 to 652)

```

AUTHORS Sato, S.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute,
The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: sato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/,
Tel: 81-438-52-3335 (ex. 2335), Fax: 81-438-52-3934)

FEATURES

Source

Location/Qualifiers
1. .652
/organism="Lotus japonicus"
/mol_type="genomic DNA"
/strain="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="Lj120K03_not"
/clone_1db="genomic TAC library"
/note="VECTOR: pYLAC7"

BASE COUNT 198 a 107 c 129 g 218 t
ORIGIN

Query Match 5.3%; Score 77.6; DB 29; Length 652;
Best Local Similarity 50.0%; Pred. No. 5.2e-10;
Matches 222; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

```

QY 212 AGGCGATTCGGTGGCGGCTATTTTGTCCGATTTTCTCTCGCGGCTTTTGGGCA 271
DB 72 ATGGATATTCCTGTCGCAATGGGATTTATGATTTTGAATTTAGAGAGTATGGAT 131
QY 272 CGGTGCGTCCGTCGCGATCATGAAGCGATTCACCGCGCCACAGCTGAACCTCCAC 331
DB 132 CCGTAGTAGTGAGAAATCAACCGGTTGTGCAATATGATACCAACCACTTTACCTC 191
QY 332 TCGTGGTCTCCCTGCTTCCGCTGTGTGGCCATTCAGAGAAACAATGACCTTTGCA 391
DB 192 TTATGTAGATGTGCACTGTGAGAGACGATGCAAGAAAGATTTGATTTGATGC 251
QY 392 TGATGTGTCCATAACCGCGGCTGTGCAAGCTCACCGGAGCGCATTTGCCCTTCG 451
DB 252 AAATGGCTAAATATCTTCTGTATATGATTAATCACTAAATAAAGTATTTCTATG 311
QY 452 TGTATTTGCGCA--TCCACAGATGGTGGCGCATGCGCTCTGTGGGCTTATCGGC 508
DB 312 TATCATATCTTAATCTCCCACTACTGTGGGATTAACAGTTTTCGCAATGTTGGGG 371
QY 509 ATCTACTTTTCCGGAACCGCGCGGCGCATAGTTTCTGGGTCCTGCGGTGAGT 568
DB 372 ATATATATTGCGCAACCAATGCTTACATTTGATTTGCGGTAATAAGATTTGAC 431
QY 569 TAACCACTGGGCGATCGCTTCCAGACGGTGTGACAGCGCGAGATTTGTGAAC 628
DB 432 AAACCTTGAATAGGCAAGTCCCGAAGTTTCAAGCGCGCAGAAATTTATTCATAGG 491
QY 629 GTGTGATGATGAATTTGTTCG 652
DB 492 GCTTGTGATTCATCTGACAC 515

```

RESULT 12
CB977199/c 617 bp mRNA linear EST 01-MAY-2003
LOCUS CAB40003 Iva_Ra_D11 Cabernet Sauvignon Berry - CAB4 Vitis vinifera
DEFINITION CDNA clone CAB40003_Iva_Ra_D11 3', mRNA sequence.
ACCESSION CB977199.1 GI:30300405
VERSION
KEYWORDS EST:
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; Vitaceae; Vitis;
1 (bases 1 to 617)
REFERENCE
AUTHORS Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook
'D'.
TITLE Expressed sequence tags from Vitis vinifera 'Cabernet Sauvignon'

JOURNAL
COMMENT unpublished
Contact: Douglas Cook, PhD
CAES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcocok@ucdavis.edu

FEATURES

Source

Seq primer: GCCAAGCATGCTGAG.
Location/Qualifiers
1. .617
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB40003_Iva_Ra_D11"
/sex="Hermaphrodite"
/dev stage="Berry on stage II, 9 mm"
/lab host="DH5alpha"
/note="Organ: Berry; Vector: pDNR; Site 1: Sfil; Site 2:
Sfil; CAB4 is a cDNA library of Vitis vinifera cv.
'Cabernet Sauvignon'. Clone 8 berries. Sampled berries were
collected from field-grown vines during stage II of berry
growth (berries were green and hard) at approximately 60
days after full bloom. The average berry size was 9
millimeters. Sampled vines were located at the University
of California, Davis, Experimental Vineyard. cDNAs were
made by oligo-dT priming and directionally cloned. 5 and
3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGACAGATGCTATTCGCGCG-3' and
5'-ATCTAGAGCGCGGCGGCGGACATG-dt(30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size selected to contain the 0.5-3 kb size fraction."

BASE COUNT 191 a 129 c 94 g 203 t
ORIGIN

Query Match 5.1%; Score 75; DB 14; Length 617;
Best Local Similarity 47.8%; Pred. No. 2.6e-09;
Matches 250; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```

QY 133 GGCTATGACAGACCTTGAAGCGGCTGAGCAAGCCAAATGCATGATGGTATT 192
DB 616 GGTATTAAGAGCGTATTTGATTTTATCAAGAAAGACAGATTAACGAGCGTCA 557
QY 193 ACTGGAAGAGCACCGTGAGGCGATCCGCTGCGCTTATTTGCGATTTTCTTC 252
DB 556 ACAAGCACAGTCAACTAAACGGTATTCGGTACAAATGGGTTATGATTTTCA 497
QY 253 CTGCGCGTCTTTGGGACAGGTCGCTCGTGCGCATCATGAAGCGATTCACCGGCC 312
DB 496 ATGGGGGTATGATGAGATCCGTAGTAGGAGAAATTCACCGTTGATGAAATG 437
QY 313 ACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCCGCTGCTGCGGATTCAG 372
DB 436 ACCAATGATTTCTACCTCTTATTTAGTGTCTTCGAGAGACAGCATGCAAGAA 377
QY 373 GACATGAGCTTTTTCATGATGATGTCATTAACCGGCGTGTGACGCGTCA 432
DB 376 GAAAGTTGAGCTTGAATGCAATGGCTAAATATCTTCGCTTTATGATTAATCA 317
QY 433 GCGCATTTGCGGTCCTGCTGCTGATTTGGCA--ATCCACAGATGGGTGGCGCAT 489
DB 316 AATAAAGATTAATTTATGATGATCAATCTTAATCACTCTACTAGTGGGGTGACA 257
QY 490 TCGGGGCTTATCTGGGCACTCACTTTTGGGAACCGGCGCGCATAGTTTCTG 549
DB 256 AGTTTGTATGTGGGAGATATCAATATGCGCAACCAATTTCTCATTTGCAAT 197
QY 550 GGTCTGCGCGGTGGGAGTTAACACATGGGCGATCGGCTTCCAGACGGTGTG 609
DB 196 GGTAAAGGTAAATTAACCAATTTGAATAAAGACAGTACTGTAAGTTTCAAGCG 137

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:57:27 ; Search time 91.3457 Seconds
(Without alignments)
7117.550 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473
Sequence: 1 GTGGAAGCGCTTCGAC.....gattgagcgttagcgag 1473

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfill1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1473	100.0	1473	4	US-09-362-899-2
2	1473	100.0	2123	4	US-09-362-899-1
3	312.6	21.2	4403765	3	US-09-103-840A-2
4	312.6	21.2	4411529	3	US-09-103-840A-1
5	142	9.6	36063	4	US-08-311-731A-140
6	116.4	7.9	1020	4	US-09-252-991A-12180
7	116.4	7.9	1338	4	US-09-252-991A-12415
8	108.6	7.4	23673	4	US-09-173-816-1
9	104.2	7.1	870	4	US-09-107-532A-170
10	97.2	6.6	444	4	US-09-252-991A-12365
11	96.2	6.5	19702	4	US-08-961-527-7
12	96.2	6.5	1830121	4	US-09-557-884-1
13	96.2	6.5	1830121	4	US-09-643-990A-1
14	93.2	6.3	2561	4	US-09-347-878-25
15	87.4	5.9	954	2	US-08-973-275-2
16	86.4	5.9	903	4	US-09-328-352-1114
17	85.6	5.8	1230025	4	US-09-198-452A-1
18	84.8	5.8	954	2	US-08-973-275-3
19	84	5.7	10303	4	US-09-634-238-410
20	79.6	5.4	999	4	US-09-252-991A-4651
21	79.6	5.4	1203	4	US-09-252-991A-4229
22	77.8	5.3	456	4	US-09-252-991A-12319
23	75.2	5.1	882	4	US-09-134-001C-966
24	72.8	4.9	2255	3	US-08-714-918-105
25	72.8	4.9	2255	3	US-09-265-315-105
26	72.8	4.9	2255	3	US-09-265-315-105
27	72.8	4.9	2255	3	US-09-266-417-105

28	54.6	3.7	3183	1	US-08-849-212-3	Sequence 3, App11
29	51	3.5	2894	1	US-08-278-091-1	Sequence 1, App11
30	51	3.5	2894	1	US-08-483-859-1	Sequence 1, App11
31	51	3.5	2894	1	US-08-472-173-1	Sequence 1, App11
32	51	3.5	2894	2	US-08-487-167-1	Sequence 1, App11
33	51	3.5	2894	2	US-08-482-816-1	Sequence 1, App11
34	51	3.5	2894	2	US-08-296-149-1	Sequence 1, App11
35	51	3.5	2894	2	US-08-801-499-1	Sequence 1, App11
36	51	3.5	2894	2	US-08-615-271-1	Sequence 1, App11
37	51	3.5	2894	3	US-09-074-660-1	Sequence 1, App11
38	51	3.5	2894	3	US-09-074-659-1	Sequence 1, App11
39	51	3.5	2894	3	US-09-106-466-1	Sequence 1, App11
40	51	3.5	2894	3	US-09-106-466A-1	Sequence 1, App11
41	51	3.5	2894	3	US-09-106-467-1	Sequence 1, App11
42	49.8	3.4	789	4	US-09-107-532A-1183	Sequence 1183, App
43	48.2	3.3	984	1	US-08-611-107-11	Sequence 11, App1
44	48.2	3.3	984	2	US-08-422-560A-11	Sequence 11, App1
45	48.2	3.3	984	3	US-08-468-793-11	Sequence 11, App1

ALIGNMENTS

```
RESULT 1
US-09-362-899-2
; Sequence 2, Application US/09362899
; Patent No. 6361986
; GENERAL INFORMATION:
; APPLICANT: Degussa-Hls AG
; APPLICANT: Forschungszentrum-Jlich GmbH
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND
; CURRENT APPLICATION NUMBER: US/09362, 899
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: DE 19924365.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1473)
; OTHER INFORMATION: accda
US-09-362-899-2

Query Match      100.0%; Score 1473; DB 4; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGGAAGCGCTTTCGACCTATGATGATGGGGGCGATGGAACACATTCAGATTGACGCTC 60
        1 GTGGAAGCGCTTTCGACCTATGATGATGGGGGCGATGGAACACATTCAGATTGACGCTC 60
DB      1 GTGGAAGCGCTTTCGACCTATGATGATGGGGGCGATGGAACACATTCAGATTGACGCTC 60
        1 GTGGAAGCGCTTTCGACCTATGATGATGGGGGCGATGGAACACATTCAGATTGACGCTC 60
QY      61 ATAGACTGCGTTTGGACCTGACAGCTTCAATTTCTTGGAATGAAATCCCAATATGAC 120
        61 ATAGACTGCGTTTGGACCTGACAGCTTCAATTTCTTGGAATGAAATCCCAATATGAC 120
DB      61 ATAGACTGCGTTTGGACCTGACAGCTTCAATTTCTTGGAATGAAATCCCAATATGAC 120
        61 ATAGACTGCGTTTGGACCTGACAGCTTCAATTTCTTGGAATGAAATCCCAATATGAC 120
QY      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGAT 180
        121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGAT 180
DB      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGAT 180
        121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGAT 180
QY      181 GAATCGTAATTAATCTGGAAGAGGACCGGTGAGGGCATTCGGTAGCCGTAATTTGTCC 240
        181 GAATCGTAATTAATCTGGAAGAGGACCGGTGAGGGCATTCGGTAGCCGTAATTTGTCC 240
DB      181 GAATCGTAATTAATCTGGAAGAGGACCGGTGAGGGCATTCGGTAGCCGTAATTTGTCC 240
        181 GAATCGTAATTAATCTGGAAGAGGACCGGTGAGGGCATTCGGTAGCCGTAATTTGTCC 240
QY      241 GATTTTCCTTCCTCGGGGCTTTTGGGACCGGTGCGGTGCGGATATGAAAGCG 300
        241 GATTTTCCTTCCTCGGGGCTTTTGGGACCGGTGCGGTGCGGATATGAAAGCG 300
DB      241 GATTTTCCTTCCTCGGGGCTTTTGGGACCGGTGCGGTGCGGATATGAAAGCG 300
        241 GATTTTCCTTCCTCGGGGCTTTTGGGACCGGTGCGGTGCGGATATGAAAGCG 300
QY      301 ATTACCGCGGACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
        301 ATTACCGCGGACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
        301 ATTACCGCGGACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
```

```
Db 301 ATTACCGCGGCACAGAGCTGAACCTCCACTGCTGCTCTCCCTGCTTCCGGTGTCG 360
Qy 361 CGCATGACAGAAAGCAATCGAGCTTTTGTGATGATGTCATTAACCGGCTGTGAG 420
Db 361 CGCATGACAGAAAGCAATCGAGCTTTTGTGATGATGTCATTAACCGGCTGTGAG 420
Qy 421 CGTCACCGCGGAGGCGCATTTTCCGCTCTGCTGATTTTGGCCATCCACAGATGGTGGC 480
Db 421 CGTCACCGCGGAGGCGCATTTTCCGCTCTGCTGATTTTGGCCATCCACAGATGGTGGC 480
Qy 481 GCCATGCGCTGTGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540
Db 481 GCCATGCGCTGTGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540
Qy 541 GGTTCCTGAGGTCTCTGCGGTGTGAGTTAAACCATGAGCATGCGCTTCCAGACGCTGTG 600
Db 541 GGTTCCTGAGGTCTCTGCGGTGTGAGTTAAACCATGAGCATGCGCTTCCAGACGCTGTG 600
Qy 601 CAGCAGCGGAGAAATTTGTGTGAAAATGTGTGATTTGATGAAATTTGTGCTCCACTGCA 660
Db 601 CAGCAGCGGAGAAATTTGTGTGAAAATGTGTGATTTGATGAAATTTGTGCTCCACTGCA 660
Qy 661 TTGCGTGCAGCGGTGCAAAAACCTCAAGTTATTCAAGCGGTAGAGCAACGATCGT 720
Db 661 TTGCGTGCAGCGGTGCAAAAACCTCAAGTTATTCAAGCGGTAGAGCAACGATCGT 720
Qy 721 TTTTCTCCAACTCTCTGCGGTGACCTTCCGCTGATGAGAGCGATTTGCGCTTCTGT 780
Db 721 TTTTCTCCAACTCTCTGCGGTGACCTTCCGCTGATGAGAGCGATTTGCGCTTCTGT 780
Qy 781 GACCCGAGAGGCGCTGAGATCGGGAGATTAATGAAAAGTTAGGGGAGACGTCGTCAAG 840
Db 781 GACCCGAGAGGCGCTGAGATCGGGAGATTAATGAAAAGTTAGGGGAGACGTCGTCAAG 840
Qy 841 CTCTCTGCTGCGCTGTGTCGCTGATGAGCCCGCTGTGCGCTTCCGCTGCGGCAATC 900
Db 841 CTCTCTGCTGCGCTGTGTCGCTGATGAGCCCGCTGTGCGCTTCCGCTGCGGCAATC 900
Qy 901 GGGGGCGGGCGCTGTGCTGATTTGGGAGATGCGCGCTTCAAGCTTTGGGCGGAGAG 960
Db 901 GGGGGCGGGCGCTGTGCTGATTTGGGAGATGCGCGCTTCAAGCTTTGGGCGGAGAG 960
Qy 961 CTGCGCTTTGGCGCTGTGAGGCAATTCGCTGGGCGGCGAGCTTAACCTGCGCATGCTGTC 1020
Db 961 CTGCGCTTTGGCGCTGTGAGGCAATTCGCTGGGCGGCGAGCTTAACCTGCGCATGCTGTC 1020
Qy 1021 ATCATCGACACTCCGCGCGCCGAATTTGTCAGAGCGGCTGAGAGCTCGGATCGCAAGC 1080
Db 1021 ATCATCGACACTCCGCGCGCCGAATTTGTCAGAGCGGCTGAGAGCTCGGATCGCAAGC 1080
Qy 1081 TCGATTGCGCGCACTTGTCCAAAGCTTATCGAGCTTCCCTCCCAACGTTTCGCTCAT 1140
Db 1081 TCGATTGCGCGCACTTGTCCAAAGCTTATCGAGCTTCCCTCCCAACGTTTCGCTCAT 1140
Qy 1141 ATTGTCAGAGGCGGTGTGGGTCGCGCGCGCTGAGCATGTCGCGCATTTGCTTACGGG 1200
Db 1141 ATTGTCAGAGGCGGTGTGGGTCGCGCGCGCTGAGCATGTCGCGCATTTGCTTACGGG 1200
Qy 1201 GCCGAAAACGCGTGTGCTGCTGATTTGTCACCAAGAGGCGCTCGGCAATCTCTTCCGC 1260
Db 1201 GCCGAAAACGCGTGTGCTGCTGATTTGTCACCAAGAGGCGCTCGGCAATCTCTTCCGC 1260
Qy 1261 GACACCAACACGCGCGCGGAAATCATAGAGCAAAAGGCTGACAGCGCATTTTAA 1320
Db 1261 GACACCAACACGCGCGCGGAAATCATAGAGCAAAAGGCTGACAGCGCATTTTAA 1320
Qy 1321 AGCCAAAGGCTTATGAGCGGAGATGATGCGCGAAACCGAGCACTTTTGAAGAAATTC 1380
Db 1321 AGCCAAAGGCTTATGAGCGGAGATGATGCGCGAAACCGAGCACTTTTGAAGAAATTC 1380
Qy 1381 GGCACAAATCAGCAAGCGCTCTCGCAATTTGATTAACAATCCGAGAGGCGGAGCGCA 1440
Db 1381 GGCACAAATCAGCAAGCGCTCTCGCAATTTGATTAACAATCCGAGAGGCGGAGCGCA 1440
```

```
Qy 1441 AGTCGCTTACACGATTTGAGCGTTTACGCGAG 1473
Db 1441 AGTCGCTTACACGATTTGAGCGTTTACGCGAG 1473

RESULT 2
US-09-362-899-1
; Sequence 1, Application US/09362899
; Patent No. 6361986
; GENERAL INFORMATION:
; APPLICANT: Degussa-Hls AG
; APPLICANT: Forschungszentrum-Jlich GmbH
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND
; CURRENT FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: DE 19924365.4
; EARLIER FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2123
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: gene
; LOCATION: (508)..(1380)
; OTHER INFORMATION: accda
US-09-362-899-1
```

```
Query Match 100.0%; Score 1473; DB 4; Length 2123;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGAGAGAGCGTTTCCGACTGATGTGTGGGCGATGAAACAACCTTACGATTGACGCTC 60
Db 508 GTGAGAGAGCGTTTCCGACTGATGTGTGGGCGATGAAACAACCTTACGATTGACGCTC 567
Qy 61 ATGACTCGGTTTGGACCTTGACAGCTTCACTTTCTTGGAAATGAATCCCAATATGAC 120
Db 568 ATGACTCGGTTTGGACCTTGACAGCTTCACTTTCTTGGAAATGAATCCCAATATGAC 627
Qy 121 AACCTCATCAAGGCTTGTGACAGAGCTTGGAGGCGGCTCGAAGCAAGGCCAAATGCGAT 180
Db 628 AACCTCATCAAGGCTTGTGACAGAGCTTGGAGGCGGCTCGAAGCAAGGCCAAATGCGAT 687
Qy 181 GAATCGTAATTAATGAGAAAGGCAACGCTGAGAGGCAATTCGCTAGCCGTTATTTGTCC 240
Db 688 GAATCGTAATTAATGAGAAAGGCAACGCTGAGAGGCAATTCGCTAGCCGTTATTTGTCC 747
Qy 241 GATTTTCTCTCTCGGCGGCTTCTTTGGGACAGGTCGCGCTGCGCATGTAAGGCG 300
Db 748 GATTTTCTCTCTCGGCGGCTTCTTTGGGACAGGTCGCGCTGCGCATGTAAGGCG 807
Qy 301 ATTACCGCGCGCAGAGAGCTGAACCTCCACTGCTGTCTCCCTGCTTCCGCTGTGCG 360
Db 808 ATTACCGCGCGCAGAGAGCTGAACCTCCACTGCTGTCTCCCTGCTTCCGCTGTGCG 867
Qy 361 CGCATGACAGAAAGCAATCGAGCTTTTGTGATGATGTCATTAACCGGCGCTGTGAG 420
Db 868 CGCATGACAGAAAGCAATCGAGCTTTTGTGATGATGTCATTAACCGGCGCTGTGAG 927
Qy 421 CGTCACCGCGGAGGCGCATTTTCCGCTCTGCTGATTTTGGCCATCCACAGATGGTGGC 480
Db 928 CGTCACCGCGGAGGCGCATTTTCCGCTCTGCTGATTTTGGCCATCCACAGATGGTGGC 987
Qy 481 GCCATGCGCTGTGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540
Db 988 GCCATGCGCTGTGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 1047
Qy 541 GGTTCCTGAGGTCTCTGCGGTGTGAGTTAAACCATGAGCATGCGCTTCCAGACGCTGTG 600
Db 1048 GGTTCCTGAGGTCTCTGCGGTGTGAGTTAAACCATGAGCATGCGCTTCCAGACGCTGTG 1107
```


601 CAGCAGCGGAGATTTGGTGAAGAACTGGTGTATGATGAAATGTTGTGCGACTCCAA 660
1108 CAGCAGCGGAGATTTGGTGAAGAACTGGTGTATGATGAAATGTTGTGCGACTCCAA 1167
661 TTGCGTGACGCGGTGGCAAAAACCTCAAGATTATTCAGCCGGTAGAGCAACGATCGT 720
1168 TTGCGTGACGCGGTGGCAAAAACCTCAAGATTATTCAGCCGGTAGAGCAACGATCGT 1227
721 TTTTCTCCAACTCTCCGCGGTGGCACTTCCGGTATGAGAGCGATTGCGGTTCTCGT 780
1228 TTTTCTCCAACTCTCCGCGGTGGCACTTCCGGTATGAGAGCGATTGCGGTTCTCGT 1287
781 GACCCGAGAGGCTTGAATTCGGGAGATTATGAGAAACGTTGGGGGCAACGATCGTCAG 840
1288 GACCCGAGAGGCTTGAATTCGGGAGATTATGAGAAACGTTGGGGGCAACGATCGTCAG 1347
841 CTTTCTGTGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 900
1348 CTTTCTGTGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1407
901 GGGGGCGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 960
1408 GGGGGCGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1467
961 CTGCGTTTGGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1020
1468 CTGCGTTTGGCGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1527
1021 ATCATGCACTCTCCGCGCGGCAATTTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1080
1528 ATCATGCACTCTCCGCGCGGCAATTTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1587
1081 TCGATTGGCGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1140
1588 TCGATTGGCGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1647
1141 ATTGCTCAAGGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1200
1648 ATTGCTCAAGGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1707
1201 GCGGAAACGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1260
1708 GCGGAAACGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1767
1261 GACCAACCACTCCGCGCGGCAATTTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1320
1768 GACCAACCACTCCGCGCGGCAATTTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1827
1321 AGCCAGGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1380
1828 AGCCAGGCGGTGGCACTTTCAGCTGAGCGGCTGTGCGGTTGCTTGGCGCGCATC 1887
1381 GGCACAACTCAGCAAGCGGCTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1440
1888 GGCACAACTCAGCAAGCGGCTGTGCAAGCGGCTGAGAGCTCGGCAATCGCAAGC 1947
1441 AGTGGCTTCAACGATTTGAGGCTTTAGCGCAG 1473
1948 AGTGGCTTCAACGATTTGAGGCTTTAGCGCAG 1980

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS

FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; TITLE OF INVENTION: represent a, t, c or g
US-09-103-840A-2

Query Match 21.2%; Score 312.6; DB 3; Length 4403765;
Best Local Similarity 54.0%; Pred. No. 1.1e-76;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

65 ACTGGTTTGGAGACCTTACAGCTTCTTCTGGAATGAAATCCCAATATGCAAC 124
1008148 ACGGGTGTGACCTCGGGATCTTTCGTAAGTATGAGAGCTGTGCGGTGCGG 1008089
125 TCAATCAAGCTATGCAAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCAATG 184
1008088 TAGCGACTCTATGCGGGAGAGCTGCGCGCTCGGCGGCGGCAACCGCGGCAAT 1008029
185 CGGTATATTCTGGAAGAGCACTGTGAGAGGCAATTCGGTATGAGCTTATTTGTCGAT 244
1008028 CGGTGCAAGCGGTGAGAGGAGCGGATTCGGGCGGCGGTGCGGTGAGT 1007969
245 TTTCTTCTGCGGGCTTCTTGGAGAGCTGCGGTGCGGTGCGGTGAGT 304
1007968 TGAATCTTCTGCGGGCTTCTTGGAGAGCTGCGGTGCGGTGAGT 1007909
305 ACCGCGCAGACAGCTGAACTCCCACTGCTGCTGCTTCCGCTGCTGCGGTGCGGCA 364
1007908 AGCGGCGCAGCGCGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1007849
365 TGCAGAAACATGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 424
1007848 TGCAGAAACATGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1007789
425 ACCGCGCAGACAGCTGAACTCCCACTGCTGCTGCTTCCGCTGCTGCGGTGCGGCA 484
1007788 ACAACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007729
485 TGGCTGCGGGGTTCACTGCGGATCTCACTTTTGGGAAACCGCGCGCGGATGAGT 544
1007728 TGGCTGCGGGGTTCACTGCGGATCTCACTTTTGGGAAACCGCGCGCGGATGAGT 1007669
545 TCTGCGGTCTGCGGTGAGTTAACCACTGCGGATGCGCTTCCAGCGGTGCGAGC 604
1007668 TCTGCGGTCTGCGGTGAGTTAACCACTGCGGATGCGCTTCCAGCGGTGCGAGC 1007609
605 AGCGGGAATTTGTGAAAACTGTGTGATGATGATGATGATGATGATGATGATGATGAT 664
1007608 CGCGGGAATTTGTGAAAACTGTGTGATGATGATGATGATGATGATGATGATGATGAT 1007549
665 GTGAGCGGTGAGCAAAACCTCAAGTTATTCAGCGGATGAGAGCAAGATGTTT 724
1007548 GACCGATGCTGATTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 1007489
725 CTCAACAATCTCTGCGGTGCA--CTTCCGATGATGAGAGGATGCGGTCTCTG 781
1007488 GCGAGAGCGCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007429
782 ACCCGAGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
1007428 GCGCGAGCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007369
842 TTTCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 901

Db 1007368 TGTCAGAACCGATCAAGGCGGAGCG---GGACACACGCTGCGCTGCGCCGCTTTG 1007312
Qy 902 GGGGCGGGCCCGTGTGTGATTTGGGCGAG-----GATGCGCGCTTCAAGC 946
Db 1007311 GCGGCAACCCAGGTGTGCTCTGGGCGAAGGCGAGTAGGCGGCGGAGGAGACTTG 1007252
Qy 947 TTGGGCGGAGAGCTGGCTTTTGGCGCTGCGATTTGCTGGCGCGGAGCTAAAC 1006
Db 1007251 TCGGGCCGCGTGTGCTTACCGGAGCCGACGCGGAGTAGGCGCTGCGCCGAGCTGTGCC 1007192
Qy 1007 TGGCGATGTGTTCATTCAGACACCTCGGCGCGGAAATTGTGCGAGCGGCTGAGAGC 1066
Db 1007191 TGCCGCTGTGTGTGCTTGTGACGCGGCGGAGCCGCGTGTGCGCGCGAGCGAAGCAGG 1007132
Qy 1067 TCGGATGCGAAGCTCGAGTTGGGCGGACCTGTTCAGCTTATCGACGCTCCCTCCGCA 1126
Db 1007131 GCGGCTGCGCGGCGAGATCGGCAATGCTGCGCGAGCTGTCACGCTGATACCCGGA 1007072
Qy 1127 CCGTTTCGATATTTATTTGTCAGGCGGCTTGGCGGCTGCGGCTGCGCTGCGCGCG 1186
Db 1007071 CCGTGTGATCTGCTGTGGGCGAGGCGAGCGGCGGCGCGGCTGCGGATTTGCGCGCG 1007012
Qy 1187 ATCTGTCTACGCGGCGGAGAAAGCGGTGCTGTGCGATTTGCGACGAGGCGGCTGCG 1246
Db 1007011 ACCGGGTCTGCGCGCTCCACCTCCAGCTGCGTGGCGCTTGTGCTCCGAGGAGCGACGG 1006952
Qy 1247 CCATCTCTTCCGCGGAGCAACACGCGCGGGAATCATAGAGGAGCAAGCGGCTGCGAG 1306
Db 1006951 CGATGTGTGCGAGCACTGCTCATGCGCGGAACTGCTGCGCGCCGCGGAGCGATCCGGT 1006892
Qy 1307 CGCAGCACTTTTAAGCCAAAGGCTTATTCAGAGGATGCTGCGCGGAAACCGAGCACTTGG 1366
Db 1006891 CGGCGCACTTGAAGTCGGGATTTGTGACACATGTGTGCGGAGTACCCCGAGCGCG 1006832
Qy 1367 TTGAAGA 1373
Db 1006831 CAGACGA 1006825

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: FRASER, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 21.2%; Score 312.6; DB 3; Length 4411529;
Best Local Similarity 54.0%; Pred. No. 1,1e-76;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

Qy 65 ACTCGTTTGGACCTGACAGCTTATTTCTTGAATGAATCTCCCAATATGACAAC 124
Db 1008147 ACGCGTGTGACACCGGGATTTTCTGTAGCTGGATAGCAGCCGCTGGGCGTGGCGG 1008088
Qy 125 TCAATCAAGGCTATGACAGACCTTGGAGCGGCTTGAAGCAAGCCAAATGCGATGAT 184
Db 1008087 TAGCGACTCTATGCGCGGAGCTTGGCGCGCTTGGGCGGCGACACCGGCGGAGCAT 1008028

Qy 185 CGGTATTTACGAGAGGACCGGTGAGGGCATTCGCTAGCCGTTATTTGTCGATT 244
Db 1008027 CGGTGACAGCCGTTGAGGAGACGCGTATTTCGGGCGGCGGCTGTGTGTGAT 1007968
Qy 245 TTTCCTTCTCTGCGGCTTCTTTTGGGACGCGTGTGCGTGTGCGCATCATGAAGCGATT 304
Db 1007967 TCGACTTCTGTGGGCGGCTGTGATTTGGGTGGACGCGCGGAACGATACCGCGCGCTG 1007908
Qy 305 ACCGCGGACAGAGCTTAAACTCCCATGTGTGTCTCCCTGCTTCCGCTGTGTGCGGCA 364
Db 1007907 AGCGGGGACCGCGAGCGGCTGCGCTACTGTGCGTACCAAGCTGCGAGGACCGCA 1007848
Qy 365 TGCAAGAAACAATTCAGCTTTTGTCTGATGAGTGTCTATTAACCGGCGGTGCGGCTG 424
Db 1007847 TGCAAGAAAGCAGCGTGTGCTTTCTGAGATGTGTGAAGATCGCTGCGGCGCATCCAGCTG 1007788
Qy 425 ACCGAGGCGCATTTTTCGCTTCTGTGTGATTTTGGCAATCCCAAGTGGTGGCGCA 484
Db 1007787 ACAACAGGCGCGCTGCTCCCTACCTGTATTTGGCGCATCCGACACGCGGTGAGATT 1007728
Qy 485 TGGCTTGTGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGGCGCATAGATT 544
Db 1007727 TCGGCTGTGGGCTTCTGTTGGGATCTCACGCTGCGGAGCGGCGGCGCTGATGCGCT 1007668
Qy 545 TCTTGGGTCTCGGCTGTGAGTTAACACTGGGCACTGGGCTTCCAGAGGCTGTGAGC 604
Db 1007667 TTTCTGSAACACGCGGTCTTATGAGTTGCTTATGCGACCCCTTCCATCCGCGCTCAA 1007608
Qy 605 AGGCGAGAAATTTGTGTAACCTGTGTGATTTGATGAAATTTGTGTGCCACTCCATTG 664
Db 1007607 CCGCGGAAATCTACGCGGCGCATGAGATCATGACGCGCTGTGCTGAGACCGGCTAC 1007548
Qy 665 GTGAGAGGTGGGAAACCTCAAGTTTATCAGCGCGTATGAGGACACGATGCTTTT 724
Db 1007547 GACCGATCTGATGCTGTGTGACGCTGTCTATGACGCTCCGAAACCTTTCGCGAC 1007488
Qy 725 CTCACAACTCTCGGCTGAGCA---CTTCCGATGATGAGAGCATTTGCGTCTGTG 781
Db 1007487 CGCAGAGCGCGCGCGCTGACCGATGTGCTCCACGTGAGGACTGCTGTGTGATGATGCGCG 1007428
Qy 782 ACCGCAAGAGCTTGAATCGGGAGATTATGAAAGTTTGGGAGGACGCTGTCAAGC 841
Db 1007427 GCGCGAGCGCGCGGCGGTGACGAGCACTGAGCAACGCGGCGACGAGCGGCTGTGT 1007368
Qy 842 TTTCTGTGGCGCTGTGCGGATTTAGCCCGGCTGTGCGGCTTCCCTGCGCGCATG 901
Db 1007367 TGTAGGAACCGATCAAGCGGAAGCG---GCGACACGCTGTGCGCTGCGCTTGG 1007311
Qy 902 GGGCGCGGCGGCTGTGATTTGGGCGAG-----GATGCGCGCTTCAAGC 946
Db 1007310 GCGGCAACCCAGGTGTGTCTCTGCGCAGCAAGGAGAGTAGGCGGCGGAGGAGCACTG 1007251
Qy 947 TTGGGCGGAGAGCTGCTTTTGGCGGTGTGCGATTTGCTGCGCGCGGAGCTAAAC 1006
Db 1007250 TCGGCGCGGCTGTGCTTACGGAACCCGACGCGGAGTAGGCGCTGCGCGGAGCTGTGCC 1007191
Qy 1007 TGGCGATGTGTTCATATGACACCTTCCGCGCGGAAATTGTGCGAGCGGCTGAGAGC 1066
Db 1007190 TGGCGTGTGTGTGATTAAGCCGCGGAGCCCGGCTGTGCGCGGCGAAGCAGG 1007131
Qy 1067 TCGGATGCGAAGCTGATTTGGCGGACCTTGTCAAGCTTATCGACGCTCCCTCCGCA 1126
Db 1007130 GCGGCGTGGCGGCAATGCGCATTTGCTTGGCGAGCTGTGATGATACCCGGA 1007071
Qy 1127 CCGTTTCGATATTTATTTGTCAGGCGTGTGCGGCTGCGGCGCATGCTGCGCGCG 1186
Db 1007070 CCGTGTGATCTGTCTGCGGCGAGGCGGCGGCGGCGGCGGCTGTGCGGCTGTGCGCGCG 1007011
Qy 1187 ATCTGTCTACGCGGCGGAAACGCGTGTGCTGTGCGATTTGCCACGAGAGGCGGCTGCG 1246
Db 1007010 ACCGGGTGTGCGCGCACTCCACGCTGTGCTGTGCGCGCTTGTGCTCCGGAAGGAGCCGCG 1006951

Query 479 GCGCATGACCTCGTGGGGTTCATCTGGGCACTCTACTTTTGGGAAACCCGGCGGAGA 538
Db 773 GCGTTTCGCGACGCTGGAGTCTGCGGACGATGCTGCGGCAACCCAGAGGCGTGA 832
Query 539 TAGGTTCTGGGCTCTCGCGGTGGTGAAGTTAACCACTGGGCACTGCTTCAGACGGTG 598
Db 833 TCGGCTTGGCGGCTCCCGCGGTGATCGAGCAGACCGTCCGGAAGAGCTGCCGAGAGCT 892
Query 599 TGCACAGCGGAGAAATTTGTGAATACTGTGTGATGATGAAATTTGTGCGCACTCC 658
Db 893 TCCAGCGTAGAGAGTTCTCTTGTAGCATGGCGCATGACATGATGATGATGATGATGATG 952
Query 659 AATTGCGTGCAGCGGTGGCAAA 680
Db 953 AGTTGCGGCGCGCTGCGCAA 974

RESULT 7
US-09-252-991A-12415/c
Sequence 12415, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12415
LENGTH: 1338
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12415

Query Match 7.9%; Score 116.4; DB 4; Length 1338;
Best Local Similarity 52.0%; Pred. No. 4,7e-23;
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

Query 179 ATGATTCGGTATTAATCTGGAAGAAGCAACCGTGGAGGCAATCCGGTACCGCTTATTTGT 238
Db 677 AGGAGCGCTGATCCCATGAGCGGCAAGCTGCAAGGATGCGGTGCTGCGCTGCGCT 618
Query 239 CCGATTTTCTCTCTCGCGCGTCTTTGGGACGCGTGGTGGCATCATGAAG 298
Db 617 TCGAGTTCTCTTCAATGAGCGGTGATGAGCGCATGCTGGCGAGCGCTTCTTCCGCG 558
Query 239 CGATTACCGCGCACAGAGTGAACCTCCACTGCTGCTCCCTGCTTCCGTTGGTG 358
Db 557 CAGCCACGTCGCCCTGAGAGCGCTGCGCTGATCTGCTTCCGCTCCGCGGCG 498
Query 359 CGCGCATGAGAAAGCAATGAGCTTTGTATGATGATGATGATGATGATGATGATGATG 418
Db 497 CGCGATGAGAGAGCGCTGATCTGCTGATGATGATGATGATGATGATGATGATGATGATG 438
Query 419 AGCGTACCGCGAGAGCGCATTTGCGTTCCTGCTGATTTTGGGCAATCCAGATGAGGTG 478
Db 437 CGCGCTGCGCGAAGAGCATCCGCTTCTGCTGATGATGATGATGATGATGATGATGATGATG 378
Query 479 GCGCATGAGCTCTGCGGCTTCACTGGGCACTCACTTTTGGGAAACCCGGCGGAGA 538
Db 377 GCGTTTCGCGACGCTGGGATGCTCGGCGAGTGTGCGGCAACCCAGAGGCGTGA 318
Query 539 TAGGTTCTGGGCTCTCGCGGTGGTGAAGTTAACCACTGGGCACTGCTTCAGACGGTG 598
Db 317 TCGGCTTGGCGGCTCCCGCGGTGATCGAGCAGACCGTCCGGAAGAGCTGCCGAGAGGT 258
Query 599 TGCACAGCGGAGAAATTTGTGAATACTGTGTGATGATGAAATTTGTGCGCACTCC 658

Db 257 TCCAGCTAGCGAGTCTCTTGTAGATGCGCGCCATGACATGATGATGATGATGATGATG 198
Query 659 AATTGCGTGCAGCGGTGGCAAA 680
Db 197 AGTTGCGGCGCGCTGCGCAA 176

RESULT 8
US-09-773-816-1/c
Sequence 1, Application US/09773816
Patent No. 6340774
GENERAL INFORMATION:
APPLICANT: Stanford University
TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
FILE REFERENCE: 28600-20210.00
CURRENT APPLICATION NUMBER: US/09/773,816
CURRENT FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: US 60/243,458
PRIOR FILING DATE: 2000-10-25
PRIOR APPLICATION NUMBER: US 60/179,305
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 1
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 23673
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(23623)
OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Query Match 7.4%; Score 108.6; DB 4; Length 23673;
Best Local Similarity 46.6%; Pred. No. 2,7e-20;
Matches 611; Conservative 0; Mismatches 619; Indels 81; Gaps 5;

Query 177 CGATGAATCGGTATTAATCTGAGAGAGCAACCGTGGAGGCAATTCGGTACCGCTTATTTT 236
Db 8617 CGAGAGGCGCTGCTCATCGAGCGGCGGAGATCCAGCGCGAGCGCTGCTGCGCGC 8558
Query 237 GTCCGATTTTCTCTCTCGCGGCTTCTTTGGGACGCGTGGCGGATCATGA 296
Db 8557 GATGACTTGGCTTCTCTCGCGGAGCTTGGGACGCGTGGCGGATCATGA 8498
Query 297 GCGCATTCACCGCGCACAGAGTGAACCTCCACTGCTGATCTCCCTGCTTCCGTTGG 356
Db 8497 CGCGCGCGAGACCGCGCTGCGCGAACAACCTGCGCTGCTGATGATGATGATGATGATG 8438
Query 357 TGGCGCATGAGAGAGCAATGAGCTTTTGTATGATGATGATGATGATGATGATGATGATG 416
Db 8437 CGCGCATGAGAGAGAGCGCGCTGCGCTGATGATGATGATGATGATGATGATGATGATGATG 8378
Query 417 GCAAGCTACCGCGAGCGCATTTGCGCTTCTGCTGATTTTGGGCAATCCAGATGAG 476
Db 8377 CGGCAACTGAGCAGAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 8318
Query 477 TGGCGCATGAGCTCTGAGGATTCATCTGGGCACTCACTTTTGGGAAACCCGGCGGCA 536
Db 8317 CGGCGTGGCGGCTCTCTTGGGACCTGTTGAGAGTATGATGATGATGATGATGATGATGATG 8258
Query 537 GATAGTTTCTTGGGCTCTGCGTGTGAGTTAACCACTGGGCACTGCGCTTCCAGACG 596
Db 8257 TCTCGGCTTCCGCGCGCGGCGGATCATGCGGCAAGACCATCCGAGAACTGCGCGCGA 8198
Query 597 TGTGACAGAGCGGAGAAATTTGTGAATACTGTGTGATGATGATGATGATGATGATGATG 656
Db 8197 GTTCCAGCGCGGAAATTAATCTCTGGAACAGCGCTGATGATGATGATGATGATGATGATG 8138
Query 657 CCAATTGGTGCAGCGGTG-----CAAAAACC 685

Db 8137 CCGCGTGGCCAGCACTGCGCCCGCTGCTGCGGGCCGACCGCAAGCCGCCAACGGCGCG 80787

QY 686 TCAAGTTATTTCAGCCGGGTAGAGGCAACGGAT-----CGTTTTCTCCAA 730

Db 8077 GGAACGGGGAACCGGCCCCGACGCGCGGCGCTTGACCGCGGGGGCGATCGCGTATGCTGTGA 8018

QY 731 CAACCTCTGGCGGTGGCACTTCCGGGTGATGAGAGCCGATTGGCGGTTCTGTGACCCGACGA 790

Db 8017 CCGCGCCCAACTCTGCGCGCGCAACGACCCGACCGCCGACGAGTGGCGCGCGCGCGCGCATCGA 7958

QY 791 --GGCGCTGAATCGGGAGGATTATGGAACGTTGGGGGGAAGCGTCAAGCTTTCTGG 848

Db 7957 CCGCGCCCAACAGCTCTGACTACATGCGCTGTGTGCAAGGACTTGTGCAACTTGGCGGG 7898

QY 849 TCGCGTCTGAGCGCATTTGAGCCCGGCTGTGGCGCTTGCCTTGACGCGCATCGGGGCGG 908

Db 7897 CGACCGCATCTCGGGCGACTGCGCCCGCATGTGCGCGGCTGGCGCGCTTGTGGGGCGG 7838

QY 909 GCCCGTGTGCTGATTGGGCA-----GGATGCGCGCTTCAAGCTTTGGGCGGACGAG 960

Db 7837 ACCGCTGCCCTTCATCGGCGCAACAGAGGGGCGACGACCGGACGAGAGCTGCGCCGCGCA 7778

QY 961 CTGGCGTTTGGGCGT-----CGTGGCATTTTGCTTGGCGCG 995

Db 7777 CTTGCGCATGCGGATCCCGTGGGCTACCGCAAGCGCGCGGATCATGCGGCTGCGGA 7718

QY 996 CGAGCTAAACCTGCGCATCGTGTCCATCATGCAACACTCGGCGCGCGAATTGTGCAAGC 1055

Db 7717 CAACCTCGGTTCTCCGGTGTGTCACTGTGTGCAACCCCGCGCTTACCCCGCGCGGT 7658

QY 1056 GCGTGAAGAGCTCGGCGATCGCAAGCTGATTGGCGGCACTTGTCCAACTTATGACGC 1115

Db 7657 CGCGAGAGGAGGGGCGAGGCGCGTGCATCGCGCAAGACTGGCGCTGATGGCGGGCT 7598

QY 1116 TCCCTCTCCCAACCGTTTGGGTTCATTATGTGTCAAGGCGTGGCGGTGGCGGCTGGCCAT 1175

Db 7597 GCGGTGCGCGGTCTGCAACCGTGTGTACCGGAGAGGCGGAGCGCGGGGCGCTTCGCGCT 7538

QY 1176 GCTCCCGCGCATCTGCTCTACAGCGGCGCGGAACCGGTGGCTGTGCGGATTGGCAACGAGA 1235

Db 7537 CGCGGTGCGCAACCGTGTCTGATGTTGAGAGACGGCACTTACTGTGATCAAGCCCGGA 7478

QY 1236 GGGCGGCTCTGGGCATCTCTTTCGCGGACCAACCAACGCGCGGGAACATATGAGCGACA 1295

Db 7477 GGGCTGCGCGGCGCATCTGTGTGAGAGAACCCGCGCGCGCGGGGTGCGCGCGCGGCGCT 7418

QY 1296 AGCGGTGCAAGCGCGCAACGCACTTTTAAGCGCAAGGCTTATGCAAGGAGTGTGCGCGAAC 1355

Db 7417 GAGGGTCAACCCCGCGCAACTCTCTGCGCTTGGGCACTGTGCAAGGCGCTCTCCCGGAAC 7358

QY 1356 CGAGCACTTTGTGAAGAAATTCTCGGCAACATAGCAACGCCCTCTCCGA 1406

Db 7357 CGAAGGCGGACCGGAGCGCAACCGGCTGCGCGCGCGCGCGCGCGCTCTCGCGGA 7307

RESULT 9

US-09-107-532A-170
Sequence 170, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:

```

1      MEDIUM TYPE: CD/ROM ISO9660
2      COMPUTER: PC
3      OPERATING SYSTEM: <Unknown>
4      SOFTWARE: ASCII
5
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/09/107,532A
8      FILING DATE: 30-Jun-1998
9
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: 60/085,598
12     FILING DATE: 14 May 1998
13
14     APPLICATION NUMBER: 60/051571
15     FILING DATE: July 2, 1997
16
17     ATTORNEY/AGENT INFORMATION:
18     NAME: Arifitello, Pamela Deneke
19     REGISTRATION NUMBER: 40,489
20     REFERENCE/DOCKET NUMBER: GTC-012
21
22     TELECOMMUNICATION INFORMATION:
23     TELEPHONE: (781) 893-5007
24     TELEFAX: (781) 893-8277
25
26     INFORMATION FOR SEQ ID NO: 170:
27     SEQUENCE CHARACTERISTICS:
28         LENGTH: 870 base pairs
29         TYPE: nucleic acid
30         STRANDEDNESS: double
31         TOPOLOGY: circular
32
33     MOLECULE TYPE: DNA (genomic)
34     HYPOTHETICAL: NO
35     ANTI-SENSE: NO
36     ORIGINAL SOURCE:
37         ORGANISM: Enterococcus faecium
38     FEATURE:
39         NAME/KEY: misc feature
40         LOCATION: (b) LOCATION 1...870
41     SEQUENCE DESCRIPTION: SEQ ID NO: 170:
42     US-09-107-532A-110

```

Query Match	7.1%	Score 104.2;	DB 4;	Length 870;
Best Local Similarity	50.9%;	Pred. No. 1.1e-19;		
Matches 247;	Conservative	0;	Mismatches 238;	Indels 0;
			Gaps	0

Qy	177	ATGAAACGGTAAATTACTGTGAGAAAGCACCGTGGAGGGCATTTCCGGTAGCCGTTATTTTGT	235
Db	323	ATGAAACGATCCTCACTGGTAAAGCAACCATTCAGGGGGATTCCCTCATTAATTGGAAGTGA	385
Qy	239	CCGATTTTCTTCTCCGCGGTTCTTTGGGACAGTTCGCTCGGTGGCATCATGAAG	295
Db	383	TGGATCCTAATTTTATTCATGGGAAGTATGGGAAGATTTGTGGAGAAAGAATCACCGAT	445
Qy	299	CGATTACCGGCCCAAGAGCTGAAATCCCATCTGCTGTCTCCCTTGCTTCCGGTGTG	355
Db	443	TATTGCAAGAACCAAAAGGAATCTTCCAGTGTCTTGTTTCACAGCTTCAGGTGCG	505
Qy	359	CGGCGATGAGAAACAAATCGAGCTTTTGTCAATAGTGTTCATTAACCGCGGCTGAC	415
Db	503	CAAGATGCAAAAGAAATCTTTTCTCTAATGCAATGCAAAAATCTGTGCTGTCTCA	565
Qy	419	AGGCTACCGCGAGCGCATTTGCCGTTCTGTGTGTAATTGCGAATCCACGATGGGTG	475
Db	563	AACGGATAGCAATAGAAAGTCTTTTATTCATCTGTTTTACGAGCCCGACGACCGAG	625
Qy	479	GAGGCATGAGCTCGTGGGGTTTCATCTGGGCAATCTCATTTTGCGGAAACCGGAGCGAGA	535
Db	623	GAGTAACTGCCAGTTTTCGCAATGGAAGGCGATTAATTCTTAGCGAAACCCCAAGCTTGA	685
Qy	539	TAGGTTTCTGGGTCTTCGCGTGTGTGAGTTAACCACTGGGCATGCGCTTTCAGACGGTG	595
Db	683	TGGGTTTTCGGAAGAGCGCTCATCGAACAAGATCAAGCAAGAACTTCCAGAGATTT	745
Qy	599	TGACAGACGCGAGAAATTGGTGA AAAATCGGTGATTTGAATTTGTGTGCGCATCC	655
Db	743	TCCAAAAGGCTGAATCTTATTAATCTCATGGTTTGTTCGATCAGATTGTTCCCGCATGG	805
Qy	659	AAATTG 663	

Db 803 AATTG 807

RESULT 10
US-09-252-991A-12365/c
Sequence 12365, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
CURRENT APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12365
LENGTH: 444
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12365

Query Match 6.6%; Score 97.2; DB 4; Length 444;
Best Local Similarity 51.6%; Pred. No. 7.3e-18;
Matches 222; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 251 TCCTCGGCGGTTCTTTGGGCGACGTCGCGTGGCATCATGAAGCGATTCACCGCG 310
Db 444 TCATGGGGGTTTCATGAGCGGCGCATGTCGGGAGCGCTTGTCGCGACCAAGTCG 385
QY 311 CCACAGAGCTGAACTCCCATGCTGATCTCCCTGCTTCGCGTGGTGGCGCATGACAG 370
Db 384 CCTGGAAGAGCGTGGCGCGGATCTGCTTCCGCTCCGCGCGCGCGCGCATGACAG 325
QY 371 AAGACATGAGCTTTTGTATGATGATGTCATTAACGCGGCTGTGACGCTCACCGCG 430
Db 324 AAGCGCTATCTGCTGATGACATGAGCGGCGCATGCTGCGCGCTGCGCGCGCG 265
QY 431 AGGCGCATTTGGCGGTCGATGATTTGGCGCATTCACGATGGTGGCGCATGAGCT 490
Db 264 AAGAAAGCATCCCGTTCGTCGATGATGACGACCGCGTCTACGCGCGCTTCCGCA 205
QY 491 CGTGGGCTTCACTCTGCGCATCTCACTTTTGCAGAACCGCGCGCGAGATAGGTTTCTGG 550
Db 204 GCGTGGCGATGCTCGGCGACGATGATGTCGCGGAAACCAAGCGCTGATCGGCTTCCGCG 145
QY 551 GTCCTGCGGTGATGAATTAACTCACTGGGCGATGCGCTTCCAGACGCTGTGCGACGAGCG 610
Db 144 GTCCCGCGCTGATGACGAGACGCGTCCGAGAACTGCGGAAAGCTTCCAGCGTAGCG 85
QY 611 AGAATTTGGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 670
Db 84 AATTCTCTCTTGAAGCATGAGCGCATGACATGATGATGATGATGATGATGATGATGATGATG 25
QY 671 CGGTGCGCAAA 680
Db 24 GCGTGGCGCA 15

RESULT 11
US-08-961-527-7/c
Sequence 7, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 19702 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-961-527-7

Query Match 6.5%; Score 96.2; DB 4; Length 19702;
Best Local Similarity 49.5%; Pred. No. 7.6e-17;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 179 ATGAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 238
Db 11553 ATGAAGCGGTTTGACAGAGAACTGCTTTATTAAGGTGACAGCTGTGGCTTGGGATTA 11494
QY 239 CCGATTTTCTTCTGCGGCTTCTTTGGGACGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGG 298
Db 11493 TGGATTTTAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11434
QY 299 CGATTAACCGGCGACAGAGTGAATCCCATGCTGATGATGATGATGATGATGATGATGATGATG 358
Db 11433 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11374
QY 359 GCGCATGACAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 418
Db 11373 CCGATGATGAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11314
QY 419 AGCGTACCGGCGAGCGCATTTGCGCTTCTGCTGATGATGATGATGATGATGATGATGATGATG 478
Db 11313 AACCGCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11254
QY 479 GCGCATGAGCGTCTGCGGCTTCACTGCGCATCTCACTTTTGCAGAACCGCGCGCGAGA 538
Db 11253 GTGTGACAGCTTCTTGTGCTATGAGAGCGATATCTTGTGATGAGAGCGCTTCCAGACGCTG 11194
QY 539 TAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598
Db 11193 TTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11134
QY 599 TCGACAGCGCGAGAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 658
Db 11133 TCGAAAGGCGAGAAATTTCTTATTAAGACATGCTTTGGATGCTATGCTCAAAAAGAGAG 11074
QY 659 AATTGCGTGCAGCGGTGCGAA 679
Db 11073 ACTTACAGATACGATTTGCTA 11053

```

RESULT 12
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match      6.5%; Score 96.2; DB 4; Length 1830121;
Best Local Similarity 49.5%; Pred. No. 5.7e-16;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 179 ATGAATCGGTAATTAATCTGGAGAGGACCGGTGAGGCGCATTCGGTAGCCGTTATTTGT 238
DB 1336980 AAGATGCGCTAATTAATTAATGACAGGTACCTTAATTAATGCAATCGTTGGCGTCGAT 1337039
QY 239 CGAATTTTCTCTTCGCGCGGTTCTTTGGGACCGGTGCGGTGCGCATCATGAAG 298
DB 1337040 CGAATTTTCTTTATGCGCGGTTCAATGGGTTCTGTAAGTGTGCAAAATTTGTTAAG 1337099
QY 299 CGATTCAACCGGCCACAGAGCTGAATCTCCACTGCTGCTCCCTGCTTCCGGTGGTG 358
DB 1337100 CGCGTGAAGAAAGCGATGGAATGATTCCTGTTGTTGTTCTCTCGAGAGTGGTGTG 1337159
QY 359 CGCGCATGACGAGAAACAAATGAGCTTTTGCATATGAGTGTCCATAACCGCGGCTGTGC 418
DB 1337160 CTCGATGACGAGAACCAATTAATTTCTTTAATGCAAAATGCAAACTAGTGCCTTAAT 1337219
QY 419 AGCGTACCGCGAGCGCATTTGCGCTTCTGCTGATTAATTTGGCAATCCCAAGATGGTG 478
DB 1337220 CTCGAATGCGTGAAGAAAGGTGTGCAATTTATTCAGTATTAACGATCGAATTTAGCG 1337279
QY 479 GCGCATGCGCTCGTGGGTTTATCTGGGCAATCTCACTTTTGGGAACCGCGCGCAGA 538
DB 1337280 GCGTATCAGCCAGTTTTCGATGATTAAGGAGATTTAAATATGCGCAGCAAAAGCCTTAA 1337339

```

```

QY 539 TAGTTTCTCGGTCCTCGCGGTGTGAGTTAACCATGGGCATCGCTTCAGACGGTG 598
DB 1337340 TTGGTTTTCAGAGGCGCAGCGTTATTTGAACAACCTGTGCGTGAATAATTCAGAAAGTT 1337399
QY 599 TGCAGCAGCGCGAGAAATTTGTGAAACTGTGTGATTTGATGAAATTTGTGCGCACTCC 658
DB 1337400 TCCAACGTATGAGTTTCTACTTGTGAGAAAGGCGCAATTTGATGATCGTGAACGTTGAG 1337459
QY 659 AATTCGTCGACCGGTGCGCA 679
DB 1337460 AATGCGTCAAACTTTAGCAA 1337480

RESULT 13
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      6.5%; Score 96.2; DB 4; Length 1830121;
Best Local Similarity 49.5%; Pred. No. 5.7e-16;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 179 ATGAATCGGTAATTAATCTGGAGAGGACCGGTGAGGCGCATTCGGTAGCCGTTATTTGT 238
DB 1336980 AAGATGCGCTAATTAATTAATGACAGGTACCTTAATTAATGCAATCGTTGGCGTCGAT 1337039
QY 239 CGAATTTTCTCTTCGCGCGGTTCTTTGGGACCGGTGCGGTGCGCATCATGAAG 298

```

Db 1337040 CGAAGCTTTCTTTATGAGGCGGTTCAATGGCTTCTGTAGTGTGCAAAATTTGTTAAAG 1337099
QY 299 CGAATTCACCGCCGACAGAGCTGAATCCCACTGCTGTCTCCCTGCTCCGGTGTG 358
Db 1337100 CGGCTGTAAGAGGAGTGAATGAAATTTGTCATTTGTGTTCTCTCGAGGTGGTGTG 1337159
QY 359 CGGCGATGAGAGAGCANTCCAGCTTTTGTGATGATGTGTCCATTAACCGGCGCTGTCC 418
Db 1337160 CTGTGTGAGAGAGGATTAATTTCTTTAATGCAAAATGCAAAACTAATGTCGTAATTG 1337219
QY 419 AGCGTACCCGCGAGGCGCATTTGCTGCTGTGATTTGGCAATCCCAAGATGGTG 478
Db 1337220 CTCGAATGAGTGAAGAGGTGTGCCATTTATTTCAATTAACGATTCGATTTAGGGG 1337279
QY 479 GCGGCATGCGCTCTGTGGGGTTCACTGGGCATCTCACTTTGGGGAACCGGCGCGCAGA 538
Db 1337280 GCGTATCACCGAGTTTGGGATGTAGGGGATTTAAATTTGCGGACCAAAAGCCTTAA 1337339
QY 539 TAGGTTTCTGGGTCCTCGCGGTGAGTTAAACCACTGGGCATGCGCTTCAGACGGTG 598
Db 1337340 TTGGTTTTCAGAGGCGCACGCTTATTTGAACAACTGTGCGTGAATAATTGCGAAGATT 1337399
QY 599 TCGACAGGCGGAGATTTGTGAATACTGTGTGATGAGAAATTTGTGCGCACTCC 658
Db 1337400 TCCAGCTGAGTGTACTTACTTGAGAAAGGGGCATGATGATGATCGTGAACCTTCAAG 1337459
QY 659 AATTGCGTACGCGGTGCCAA 679
Db 1337460 AATGCGTCAAACTTTAGCAA 1337480

RESULT 14

US-09-347-878-25
; Sequence 25, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1131)..(2399)
; OTHER INFORMATION: Escherichia coli nucleic acid encoding
; OTHER INFORMATION: folypolyglutamate synthetase-dihydrofolate
; OTHER INFORMATION: synthetase
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: M32445/Genbank
US-09-347-878-25

Query Match 6.3%; Score 93.2; DB 4; Length 2561;
Best Local Similarity 47.9%; Pred. No. 2,1e-16;
Matches 269; Conservative 0; Mismatches 293; Indels 0; Gaps 0;
QY 200 AAGGCAAGCGTGAAGGCAATCCGGTATTTTGTCCGATTTTCTTCTCGGCG 259
Db 481 AAGGCAAGCTGTATGAAAGCCGGTGTGCTGCGCAATTCAGTTTCGCTTATGAGG 540
QY 260 GTTCTTTGGGCAAGGCGCGGTGCGGATCATGAAGGCAATTCAGCGCGCACAGGC 319
Db 541 GTTCAATGAGGCTCTGTTGTGGGTGACAGTTTGTGTGTCGCTTACGAGGCGCTGGAAG 600
QY 330 TGAACCTCCCACTGCTGTCTCCCTGTCTTCGATGTGTGCGGCATGACGAGAACAATC 379
Db 601 ATTAAGTCCCGCTGATCTCTTCTCGGCTCTGTGTGGCGCAGTATGAGAGAAAGACTGA 660

QY 380 GAGCTTTGTATGATGATGTCTCATTAACCGGCGCTGTGCAAGCTTACCGGAGCGCAT 439
Db 661 TGTGCTGATGACAGATGGCGAAAACTCTGTGGGACATGGCAAAAATGACAGAGCGGCT 720
QY 440 TGGCGTCTCGTGTATTTGGCAATCCCAAGATGGGTGGCGGCATGAGCTCGTGGGGTT 499
Db 721 TCCGTTACATCTCCGTGTGACCAACCGAGATGGGCGGCTGTTCTTGTCAAGTTTCGCA 780
QY 500 CATCTGGGCAATCTCACTTTTGGGAAACCGGCGCGAGATAGTTTCTTGGGTCTGCG 559
Db 781 TGTGGGCGATCTCAACATCGCTGAACCGAAAGGTTAATGCTTTCGCTCGCGCTGT 840
QY 560 TGTGTGACTTAAACACATGGGATGCGCTTCCAGACGATGTGACACAGGCGAGAAATTTGG 619
Db 841 TATGCAAGACGCTGTGCGGAAAAAATGCGCGCTGTGATTCAGCGAGATGAATTCCTGA 900
QY 620 TGAACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
Db 901 TCGAAGAAAGGCGCATGACATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 680 AAACCTCAAGTTATTCAGCGGTGAGGCAAGCATGCTTTTCTTCAACAACTCTCTG 739
Db 961 GCATTCGCGCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
QY 740 GCGTGCACCTTCGGTATGGA 761
Db 1021 TAGGTGTACCCCGGTACCGGA 1042

RESULT 15

US-08-973-275-2
; Sequence 2, Application US/08973275B
; Patent No. 5958706
; GENERAL INFORMATION:
; APPLICANT: MATSUNAGA, Tadashi
; APPLICANT: KAWIYA, Shinji
; APPLICANT: NAKA, Kenryo
; TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS
; TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE
; FILE REFERENCE: MATSUNAGA
; CURRENT APPLICATION NUMBER: US/08/973,275B
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: PCT/JP97/01043
; EARLIER FILING DATE: 1997-03-27
; EARLIER APPLICATION NUMBER: JP 8-97536
; EARLIER FILING DATE: 1996-03-28
; EARLIER APPLICATION NUMBER: JP 8-146833
; EARLIER FILING DATE: 1996-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Magnetospirillum AMB-1
; FEATURE:
; NAME/KEY: gene
; LOCATION: (217)..(702)
; OTHER INFORMATION: Predicted region encoding membrane bound portion
; OTHER INFORMATION: of the mps gene product. "n" is any of a, t, c or
; OTHER INFORMATION: g.
US-08-973-275-2

Query Match 5.9%; Score 87.4; DB 2; Length 954;
Best Local Similarity 33.6%; Pred. No. 5.8e-15;
Matches 152; Conservative 54; Mismatches 247; Indels 0; Gaps 0;
QY 951 GCGGAGAGAGTGGCTTTTGGCGGTGCGATTTTCGTCGCGCGAGCTAACTGCC 1010
Db 393 RCGAGAGGATYAGNABRCNABRMGNATNAGANATGACGACGATGATGATGATGATG 452
QY 1011 GATGTGTCAATATGACACCTTCGCGCGGCAATGTGTGACGCGGCTAGAGAGCTCGG 1070

Db 453 NATHTACNNTGTGTGAACGCGNGGNCNTAYCCNGGNTGTGAAGCNGARGCNMGNGG 512
QY 1071 CATGCCAAGCTGATTTGGCGGCACCTTGCCAAAGCTTATCGAGCGTCCCTCCCAACCGT 1130
Db 513 NCARGCNGARGCNATHGCGNMGNSNATHGARAONTGYTMAAYGTMMGNTGCCNTYNTGT 572
QY 1131 TTCGGTCATTTATTTGGTCAGAGGCGCTTGGCGGCGCTGGCCATGCTGCCGCCGATCT 1190
Db 573 NMSGNTNATHATHGNGARGNGGNGWSNGNGNGCNAHGCNTYNTGNCNACNGNAAYAC 632
QY 1191 GGTCTACGCGGCGGAAAACGCGTGGCTGTCCGCATTGCCACCAAGAGGCGCCTCGGCCAT 1250
Db 633 NGTNTNATGYTNGARCAVGCNATHATAYVMSGNTNATHMSGNCNGARGGNTGYGCNWSNAT 692
QY 1251 CCTCTCGCGGACACCAACCGCGCGGAAATCATAGCGGACCAAGCGCTGACAGCGCA 1310
Db 693 HTTNTGCGMGNSNGCNAARAYGCNGCNGARCARYTMMGNTNACNGCNCA 752
QY 1311 CGCACTTTTAAAGCCAAAGGCTTATCGACGGGATCGTCCGGAACCGAGCACTTTGTTGA 1370
Db 753 RGAYYTNCAYARNTNMSGNATHATHGATVWSNGTNGTNCNGARCCNATGGGNGGNCNCA 812
QY 1371 AGAAATTTCTCGGCACAATCAGCAACGCCCTCTC 1403
Db 813 YMGNAAYCCNGAYYTATGATGCARACNTYNTWS 845

Search completed: November 13, 2003, 00:15:17
Job time : 124.346 secs

THIS PAGE BLANK (USPTO)

Db 1 GTGAGAAAGCGTTTTCCGACTATGATGATGATGAGGAGCAACACTTCAGACTTGAAGCTTC 60
QY ATGAGCTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAAATGAACCTCCCAATATGAC 120
Db 61 ATGAGCTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAAATGAACCTCCCAATATGAC 120
QY 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGGAGGAGCCCAATGCGAT 180
Db 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGGAGGAGCCCAATGCGAT 180
QY 181 GAATGGTAAATTAATCTGAGAGAAAGGCACTGGAGGGGCAATTCGGTATGCGCTTATTTTGTC 240
Db 181 GAATGGTAAATTAATCTGAGAGAAAGGCACTGGAGGGGCAATTCGGTATGCGCTTATTTTGTC 240
QY 241 GATTTTCTCTGCGGCGGTTCTTTGGGCAAGCTGCGGTGGTGGGATCAATGAAGGCG 300
Db 241 GATTTTCTCTGCGGCGGTTCTTTGGGCAAGCTGCGGTGGTGGGATCAATGAAGGCG 300
QY 301 ATTCAACCGGCCACAGAGCTGAACCTCCCACTGCTGATCTCCCTGCTTCCGGTGGTGG 360
Db 301 ATTCAACCGGCCACAGAGCTGAACCTCCCACTGCTGATCTCCCTGCTTCCGGTGGTGG 360
QY 361 CGCATGCAAGAAAGCAATGAGCTTTTGTATGATGATGATGATGATGATGATGATGATGAT 420
Db 361 CGCATGCAAGAAAGCAATGAGCTTTTGTATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 CGTCAACCGGCCAGGCGCATTTGCGGTTCCGATGATGATGATGATGATGATGATGATGATGAT 480
Db 421 CGTCAACCGGCCAGGCGCATTTGCGGTTCCGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 GGCATGAGCTGCTGAGGCTTCACTGAGGCACTCACTTTTGGAGAACCGGCGCGAGATA 540
Db 481 GGCATGAGCTGCTGAGGCTTCACTGAGGCACTCACTTTTGGAGAACCGGCGCGAGATA 540
QY 541 GGTTCCTGAGCTGCTGAGGCTTCACTGAGGCACTCACTTTTGGAGAACCGGCGCGAGATA 600
Db 541 GGTTCCTGAGCTGCTGAGGCTTCACTGAGGCACTCACTTTTGGAGAACCGGCGCGAGATA 600
QY 601 CAGCAGGCGGAGAAATTTGGTGAACCTGGTGAATGATGATGATGATGATGATGATGATGATGAT 660
Db 601 CAGCAGGCGGAGAAATTTGGTGAACCTGGTGAATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 TTGCGTGAAGCGGTGGCAAAAACCTTCAGAGTTATTCAGCCGGTGAAGGCAACGATGCT 720
Db 661 TTGCGTGAAGCGGTGGCAAAAACCTTCAGAGTTATTCAGCCGGTGAAGGCAACGATGCT 720
QY 721 TTTTCTCCAACTCTGAGGCTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 721 TTTTCTCCAACTCTGAGGCTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 GACCCGCAAGGCTGGAATCGGGAGATTAATGAAACGTTGGGGGCAAGAGCTGCTCAAG 840
Db 781 GACCCGCAAGGCTGGAATCGGGAGATTAATGAAACGTTGGGGGCAAGAGCTGCTCAAG 840
QY 841 CTTTCTGAGCGCTGAGCGCATTCAGCCGCTGATGAGCGGCTTCCGCTGAGCGGCAATC 900
Db 841 CTTTCTGAGCGCTGAGCGCATTCAGCCGCTGATGAGCGGCTTCCGCTGAGCGGCAATC 900
QY 901 GGGGGGCGGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 901 GGGGGGCGGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 CTGCGTTTGGGCGCTGAGGCAATTCGATGAGCGGCGGAGCTGAACCTGCGCATGCTGCTC 1020
Db 961 CTGCGTTTGGGCGCTGAGGCAATTCGATGAGCGGCGGAGCTGAACCTGCGCATGCTGCTC 1020
QY 1021 ATCATGCAACCTCCGCGCGCATTCGATGAGCGGCGGAGAGCTGAGAGCTGAGAGCTGAGAG 1080
Db 1021 ATCATGCAACCTCCGCGCGCATTCGATGAGCGGCGGAGAGCTGAGAGCTGAGAGCTGAGAG 1080
QY 1081 TCGATTTGGCGGCACTTTGTCAGAGCTTATGAGCGCTCCCTCCCAACCGGTTTGGTCAAT 1140
Db 1081 TCGATTTGGCGGCACTTTGTCAGAGCTTATGAGCGCTCCCTCCCAACCGGTTTGGTCAAT 1140

QY 1141 ATGTCAGAGGCGTTTGGCGGATGAGCGCTGAGGCAATGCTGACCGGCAATGCTGATGACGG 1200
Db 1141 ATGTCAGAGGCGTTTGGCGGATGAGCGCTGAGGCAATGCTGACCGGCAATGCTGATGACGG 1200
QY 1201 GCCGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1201 GCCGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 GACACCAACCGGCGGAGAAATCATGAGCAAGCGCTGAGGCGGCACTTTTA 1320
Db 1261 GACACCAACCGGCGGAGAAATCATGAGCAAGCGCTGAGGCGGCACTTTTA 1320
QY 1321 AGCCAAAGGCTTATGCAAGGAGATGCTGCGGAAACCGAGACCTTTTGAAGAAATTC 1380
Db 1321 AGCCAAAGGCTTATGCAAGGAGATGCTGCGGAAACCGAGACCTTTTGAAGAAATTC 1380
QY 1381 GGCACATCAAGCAAGCGCTCTCGAATGATGATGATGATGATGATGATGATGATGATGAT 1440
Db 1381 GGCACATCAAGCAAGCGCTCTCGAATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1473
Db 1441 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1473

RESULT 2
US-10-024-370-2
; Sequence 2, Application US/10024370
; Publication No. US20020142405A1
; GENERAL INFORMATION:
; APPLICANT: TILG, YVONNE
; APPLICANT: ELMANN, BERND
; APPLICANT: EGGELING, LOTHAR
; APPLICANT: SAHM, HERMANN
; APPLICANT: MOCKEL, BETTINA
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
; TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
; FILE REFERENCE: 21123-284139-MAS
; CURRENT FILING DATE: US/10/024,370
; PRIOR APPLICATION NUMBER: 09/362,899
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: DE 199 24 365.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1473)
; OTHER INFORMATION: accda
US-10-024-370-2

Query Match 100.0%; Score 1473; DB 13; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAAAGCGTTTTCCGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 1 GTGAGAAAGCGTTTTCCGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
QY 61 ATGAGCTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAAATGAACCTCCCAATATGAC 120
Db 61 ATGAGCTGGTTTGGAGCCCTGACAGCTTCATTTCTTGGAAATGAACCTCCCAATATGAC 120
QY 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGGAGGAGCCCAATGCGAT 180
Db 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGGAGGAGCCCAATGCGAT 180

QY 181 GAATCGTAATTAATGAGAGGACCGTGAAGGATTCGGTAGCCGTTATTTGTCC 240
 DB 181 GATCGTAATTAATGAGAGGACCGTGAAGGATTCGGTAGCCGTTATTTGTCC 240
 QY 241 GATTTTCTTCTCTCGCGGTTCTTTGGGACCGTGCCTCGGTGGCATCATGAAGGC 300
 DB 241 GATTTTCTTCTCTCGCGGTTCTTTGGGACCGTGCCTCGGTGGCATCATGAAGGC 300
 QY 301 ATTCAACCGGACAGAGGTGAACCTCCAGTGTGTCTCCCGTCTCCGTTGGTGG 360
 DB 301 ATTCAACCGGACAGAGGTGAACCTCCAGTGTGTCTCCCGTCTCCGTTGGTGG 360
 QY 361 CCAGTGAAGAGACATGAGCTTTTGCATGATGTGTCCATACCGCGCTGTGCAG 420
 DB 361 CCAGTGAAGAGACATGAGCTTTTGCATGATGTGTCCATACCGCGCTGTGCAG 420
 QY 421 CGTCAACCGGACAGAGGTGAACCTCCAGTGTGTCTCCCGTCTCCGTTGGTGG 480
 DB 421 CGTCAACCGGACAGAGGTGAACCTCCAGTGTGTCTCCCGTCTCCGTTGGTGG 480
 QY 481 GGCATGAGCTCGTGGGTTCACTGGGCACTCACTTTTGGGAACCGCGCGGAGAT 540
 DB 481 GGCATGAGCTCGTGGGTTCACTGGGCACTCACTTTTGGGAACCGCGCGGAGAT 540
 QY 541 GGTTCCTGAGTCTTGGCGTGGAGTTAACCACTGGGATGCGCTTCAGACGGTGG 600
 DB 541 GGTTCCTGAGTCTTGGCGTGGAGTTAACCACTGGGATGCGCTTCAGACGGTGG 600
 QY 601 CAGCAGCGGAGAAATTTGTGTGAACCTGTGTGATGATGAATTTGTGTGCACTCA 660
 DB 601 CAGCAGCGGAGAAATTTGTGTGAACCTGTGTGATGATGAATTTGTGTGCACTCA 660
 QY 661 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTATTCACCGGTGAAGCAACGATCGT 720
 DB 661 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTATTCACCGGTGAAGCAACGATCGT 720
 QY 721 TTTTCTCCAACTCTCTGGCGTGGCACTTCGCGTGAAGGAGGATTCGCGTCTGT 780
 DB 721 TTTTCTCCAACTCTCTGGCGTGGCACTTCGCGTGAAGGAGGATTCGCGTCTGT 780
 QY 781 GACCCGAGAGGCTTGAATTCGGGAGATTAATGAACCTGTGGGAGAGAGCTGTCA 840
 DB 781 GACCCGAGAGGCTTGAATTCGGGAGATTAATGAACCTGTGGGAGAGAGCTGTCA 840
 QY 841 CTTTCTGAGCGCGTGGCGCATTTGAGACCGCGCTGTGCGGTTGCTGTGCGCAT 900
 DB 841 CTTTCTGAGCGCGTGGCGCATTTGAGACCGCGCTGTGCGGTTGCTGTGCGCAT 900
 QY 901 GGGGAGCGGCGCTGTGATGATGAGGAGATTCGCGCTTCAGCGTTGGGCGGAGAG 960
 DB 901 GGGGAGCGGCGCTGTGATGATGAGGAGATTCGCGCTTCAGCGTTGGGCGGAGAG 960
 QY 961 CTGCGTTTTCGCGTGTGAGCAATTCGCTGAGCGCGGAGACTGAACCTGCGCAT 1020
 DB 961 CTGCGTTTTCGCGTGTGAGCAATTCGCTGAGCGCGGAGACTGAACCTGCGCAT 1020
 QY 1021 ATTCATGACACTCCGCGCGCGCAATTTGTGAGGCGGCTGAGAGCTGCGCATTCGA 1080
 DB 1021 ATTCATGACACTCCGCGCGCGCAATTTGTGAGGCGGCTGAGAGCTGCGCATTCGA 1080
 QY 1081 TCGATTGCGCGCACTTGTCCAAAGCTTAATGACGCTCCCTCCCAACGTTTCGGTCA 1140
 DB 1081 TCGATTGCGCGCACTTGTCCAAAGCTTAATGACGCTCCCTCCCAACGTTTCGGTCA 1140
 QY 1141 ATTGATCAAGGCGTTGAGCGGCGGCTGAGCATGCTGCGCGCATCTGATCAAGCG 1200
 DB 1141 ATTGATCAAGGCGTTGAGCGGCGGCTGAGCATGCTGCGCGCATCTGATCAAGCG 1200
 QY 1201 GCGGAAACGCGTGTGCTGCGCATTTGCAACGAGGCGCTCTGCGCATCTCTTCCGC 1260
 DB 1201 GCGGAAACGCGTGTGCTGCGCATTTGCAACGAGGCGCTCTGCGCATCTCTTCCGC 1260

QY 1261 GACACCAACCAACCGCGGAAATCATAGAGCAGACGAGCGTGCAGCGCACCACTTTA 1320
 DB 1261 GACACCAACCAACCGCGGAAATCATAGAGCAGACGAGCGTGCAGCGCACCACTTTA 1320
 QY 1321 ACCCAAGGCTTATGACGCGGATGCTGCGCGAAACCGAGCACTTTGTGAATAATCTC 1380
 DB 1321 ACCCAAGGCTTATGACGCGGATGCTGCGCGAAACCGAGCACTTTGTGAATAATCTC 1380
 QY 1381 GGCACAAATCAGAACGCCCTCTCCGAATTTGATCAATCCGAGAGAGGCGGAGCGCAC 1440
 DB 1381 GGCACAAATCAGAACGCCCTCTCCGAATTTGATCAATCCGAGAGAGGCGGAGCGCAC 1440
 QY 1441 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1473
 DB 1441 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1473

RESULT 3
 US-10-024-370-1
 ; Sequence 1, Application us/10024370
 ; Publication No. US20020142405A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TILG, YVONNE
 ; APPLICANT: ELKMANN, BERND
 ; APPLICANT: EGGELING, LOTHAR
 ; APPLICANT: SAHM, HERMANN
 ; APPLICANT: MCKEL, BETTINA
 ; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
 ; TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
 ; FILE REFERENCE: 21123-284139-MAS
 ; CURRENT APPLICATION NUMBER: US/10/024,370
 ; PRIOR FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/362,899
 ; PRIOR FILING DATE: 1999-07-29
 ; PRIOR APPLICATION NUMBER: DE 199 24 365.4
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 2123
 ; TYPE: DNA
 ; ORGANISM: Corynebacterium glutamicum
 ; FEATURE:
 ; NAME/KEY: gene
 ; LOCATION: (508)..(1980)
 ; OTHER INFORMATION: accda
 ; US-10-024-370-1

Query Match 100.0%; Score 1473; DB 13; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAGCGTTTCCGACTATGTGTGGGAGATGAAACACTTCAGATTGACGCTC 60
 DB 508 GTGAGAGCGTTTCCGACTATGTGTGGGAGATGAAACACTTCAGATTGACGCTC 60
 QY 61 ATGACTCGTTTGGAGCTTGAACGCTTCAATTTCTTGAATGAATCTCCCAATATGAC 120
 DB 568 ATGACTCGTTTGGAGCTTGAACGCTTCAATTTCTTGAATGAATCTCCCAATATGAC 120
 QY 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGGCGGCTGGAACCAAGGCCAATTCGAT 180
 DB 628 AACCTCAATCAAGGCTATGACAGACCTTGGAGGCGGCTGGAACCAAGGCCAATTCGAT 180
 QY 628 AACCTCAATCAAGGCTATGACAGACCTTGGAGGCGGCTGGAACCAAGGCCAATTCGAT 180
 DB 688 GAATCGTAATTAATCGAGAGAGGACCGTGAAGGAGCATTCGGTAGCCGTTATTTGTCC 240
 QY 241 GATTTTCTTCTCTGCGGCTTCTTTGGGACCGTGCCTGCGTGGCATCATGAAGGC 300
 DB 748 GATTTTCTTCTCTGCGGCTTCTTTGGGACCGTGCCTGCGTGGCATCATGAAGGC 300

QY 301 ATTCACGCGGCGCAGAGAGTGAATCTCCACTGCTGCTCTCCCTGCTTCCGCTGCTGTCG 360
Db 808 ATTACCGCGCGCAGAGAGTGAATCTCCACTGCTGCTCTCCCTGCTTCCGCTGCTGTCG 867
QY 361 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGCTGTCATTAACCGCGCTGTGCG 420
Db 868 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGCTGTCATTAACCGCGCTGTGCG 927
QY 421 CGTACCGCGGAGGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Db 928 CGTACCGCGGAGGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 987
QY 481 GGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 988 GGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1047
QY 541 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 1048 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1107
QY 601 CAGCAGCGGAGGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 1108 CAGCAGCGGAGGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1167
QY 661 TTGCGTGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 1168 TTGCGTGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1227
QY 721 TTTTCTCCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 1228 TTTTCTCCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1287
QY 781 GACCGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 1288 GACCGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1347
QY 841 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 1348 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
QY 901 GGGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 1408 GGGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1467
QY 961 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 1468 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
QY 1021 ATCATCGACACCTCGCGCGCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1528 ATCATCGACACCTCGCGCGCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1587
QY 1081 TCGATTTGCGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1588 TCGATTTGCGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1647
QY 1141 ATTGGTCAAGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 1648 ATTGGTCAAGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1707
QY 1201 GCCGAAAACGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1708 GCCGAAAACGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1767
QY 1261 GACACCAACACGCGCGGAAATCATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
Db 1768 GACACCAACACGCGCGGAAATCATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1827
QY 1321 AGCGAAGGCGCTTATGAGAGGAGTGTGCGCGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
Db 1828 AGCGAAGGCGCTTATGAGAGGAGTGTGCGCGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1887
QY 1381 GGCACAAATCAGCAAGCGCCTCTCCGAAATTTGATTAACAATCCGAGAGGCGGAGCGGAGC 1440

Db 1888 GGCACAAATCAGCAAGCGCCTCTCCGAAATTTGATTAACAATCCGAGAGGCGGAGCGGAGC 1947
QY 1441 AGTGGCTTCAACAGATTTGAGCGTTTACGCCAG 1473
Db 1948 AGTGGCTTCAACAGATTTGAGCGTTTACGCCAG 1980
RESULT 4
US-09-738-626-1/c
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
Query Match 100.0%; Score 1473; DB 10; Length 3309400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGAGAGAGCGTTTCCGACTATGCTGTGGGCGATGGAACAACCTTAGATGAGCTC 60
Db 881114 GTGAGAGAGCGTTTCCGACTATGCTGTGGGCGATGGAACAACCTTAGATGAGCTC 881055
QY 61 ATGACTCGGTTTGAACCTGACAGCTTCAATTTCTTGAATGAATCTCCCAATATGAC 120
Db 881054 ATGACTCGGTTTGAACCTGACAGCTTCAATTTCTTGAATGAATCTCCCAATATGAC 880995
QY 121 AACCTCAATCAAGCTATGACAGACCTTGAGCGGCTGGAAGCAAGGCAATGCGAT 180
Db 880994 AACCTCAATCAAGCTATGACAGACCTTGAGCGGCTGGAAGCAAGGCAATGCGAT 880935
QY 181 GAATCGGTATTAATGAGAAAGGACCGTGAAGGCAATTCGGTACCGCTTATTTTGTCC 240
Db 880934 GAATCGGTATTAATGAGAAAGGACCGTGAAGGCAATTCGGTACCGCTTATTTTGTCC 880875
QY 241 GATTTTCTTCTCGCGCGGCTTCTTGGGAGCGAGTGCAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 880874 GATTTTCTTCTCGCGCGGCTTCTTGGGAGCGAGTGCAGGAGGAGGAGGAGGAGGAGGAGG 880815
QY 301 ATTCACCGCGCACAAGAGTGAATCTCCACTGCTGCTCTCCCTGCTTCCGCTGCTGTCG 360
Db 880814 ATTCACCGCGCACAAGAGTGAATCTCCACTGCTGCTCTCCCTGCTTCCGCTGCTGTCG 880755
QY 361 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGCTGTCATTAACCGCGCTGTGCG 420
Db 880754 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGCTGTCATTAACCGCGCTGTGCG 880695

QY 421 GGTCAACCGGAGGCGCATTGGCTTCTGATGATTTGCGCATCCAGATGGGTGC 480
 DB 880634 GGTCAACCGGAGGCGCATTGGCTTCTGATGATTTGCGCATCCAGATGGGTGC 880635
 QY 481 GCGATGGCTCTGGGGGTTCAATGCGGATCTCACTTTTGGGAAACCGGGCGGAGAT 540
 DB 880634 GCGATGGCTCTGGGGGTTCAATGCGGATCTCACTTTTGGGAAACCGGGCGGAGAT 880635
 QY 541 GGTTCCTGGGCTCTGGGAGTGAACCACTGGGATGGCGCTTCAGACGGTGG 600
 DB 880574 GGTTCCTGGGCTCTGGGAGTGAACCACTGGGATGGCGCTTCAGACGGTGG 880575
 QY 601 CAGCAGCGGAGAAATTTGGTGAACCTGATGATTTGATGATGATTTGCTGCTCA 660
 DB 880514 CAGCAGCGGAGAAATTTGGTGAACCTGATGATTTGATGATGATTTGCTGCTCA 880515
 QY 661 TTGCGTGAAGCGGTGGCAAAAACCTCAAGTTATTACCGGTGAGGCAACGATCGT 720
 DB 880454 TTGCGTGAAGCGGTGGCAAAAACCTCAAGTTATTACCGGTGAGGCAACGATCGT 880455
 QY 721 TTTTCCAACTCTGGGCTGGGACCTTGGGATGAGGAGGCAATGGCGGTTCTGCT 780
 DB 880394 TTTTCCAACTCTGGGCTGGGACCTTGGGATGAGGAGGCAATGGCGGTTCTGCT 880395
 QY 781 GACCCGAGAGGCTGGAATCGGAGATTAATGAAACGTTGGGGCAGACGTCGTCAG 840
 DB 880334 GACCCGAGAGGCTGGAATCGGAGATTAATGAAACGTTGGGGCAGACGTCGTCAG 880335
 QY 841 CTTTCTGGTGGCGGTGCTGGCGCATTTAGACCCGGCTGTGGCGGTTGCCGCGCATC 900
 DB 880274 CTTTCTGGTGGCGGTGCTGGCGCATTTAGACCCGGCTGTGGCGGTTGCCGCGCATC 880275
 QY 901 GGGGGCGGGCGGCTGTGATGAGGAGATTCGCCCTTACCGCTTGGGCGGAGAG 960
 DB 880214 GGGGGCGGGCGGCTGTGATGAGGAGATTCGCCCTTACCGCTTGGGCGGAGAG 880215
 QY 961 CTGCGTTTGGCGGTGCTGGCATTTTCGCTGGCGCGGAGCTAAACCTGCGCATCTGTC 1020
 DB 880154 CTGCGTTTGGCGGTGCTGGCATTTTCGCTGGCGCGGAGCTAAACCTGCGCATCTGTC 880155
 QY 1021 ATCATGCACTCTCGGCGCGGAGATTTGTCAGGCGGCTGAGAGCTCGGATTCGCAAGC 1080
 DB 880094 ATCATGCACTCTCGGCGCGGAGATTTGTCAGGCGGCTGAGAGCTCGGATTCGCAAGC 880095
 QY 1081 TCGATGAGGCGACCTTGTCAAGCTTATCGAGCTCCCTCCCAACCGTTCCGTCATT 1140
 DB 880034 TCGATGAGGCGACCTTGTCAAGCTTATCGAGCTCCCTCCCAACCGTTCCGTCATT 880035
 QY 1141 ATTGTCAGGCGGTGCGGTGGCGGCTGAGCTGATGCTGCGCATCTGATTCAGCG 1200
 DB 87974 ATTGTCAGGCGGTGCGGTGGCGGCTGAGCTGATGCTGCGCATCTGATTCAGCG 87975
 QY 1201 GCCGAAACGCGGTGCTGCGCATTTGCCACGAGGGGCGCTCGGCAATCTCTTCGCG 1260
 DB 87914 GCCGAAACGCGGTGCTGCGCATTTGCCACGAGGGGCGCTCGGCAATCTCTTCGCG 87915
 QY 1261 GACACCAACGCGCGGAGAAATCATAGAGGAGCAAGGCGTGCAGGCGGCACTTTTA 1320
 DB 879854 GACACCAACGCGCGGAGAAATCATAGAGGAGCAAGGCGTGCAGGCGGCACTTTTA 879855
 QY 1321 AGCCAAAGGCTTATGAGCGGATCTGCGCGGAAACCGAGCACTTTTGAAGAAATTC 1380
 DB 879794 AGCCAAAGGCTTATGAGCGGATCTGCGCGGAAACCGAGCACTTTTGAAGAAATTC 879795
 QY 1381 GGCACATTCAGCAAGCGCTCTCGAATTTGATTAACAATTCGAGAGGCGGAGCGGAC 1440
 DB 879734 GGCACATTCAGCAAGCGCTCTCGAATTTGATTAACAATTCGAGAGGCGGAGCGGAC 879735
 QY 1441 AGTGGCTTCAACGATTTGAGGCTTTAGGCGAG 1473
 DB 879674 AGTGGCTTCAACGATTTGAGGCTTTAGGCGAG 879675

RESULT 5
 US-09-712-363-35
 ; Sequence 35, Application US/09712363
 ; Patent No. US20020164588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eisenberg, David
 ; APPLICANT: Rotstein, Sergio H.
 ; APPLICANT: Marcotte, Edward M.
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 ; FILE REFERENCE: 07419-032001
 ; CURRENT APPLICATION NUMBER: US/09/712,363
 ; PRIOR FILING DATE: 2000-11-13
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246
 ; PRIOR FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/179,531
 ; PRIOR FILING DATE: 2000-02-01
 ; PRIOR APPLICATION NUMBER: 60/117,844
 ; PRIOR FILING DATE: 1999-01-29
 ; PRIOR APPLICATION NUMBER: 60/118,206,
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: 60/126,593
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 60/134,093
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/134,092
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/165,124
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: 60/165,086
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 292
 ; SOFTWARE: PastSeq for Windows Version 4.0
 ; SEQ ID NO 35
 ; LENGTH: 1488
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-712-363-35
 Query Match 21.2%; Score 312.6; DB 10; Length 1488;
 Best Local Similarity 54.0%; Pred. No. 5.5e-87;
 Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
 QY 65 ACTGGTTTGGACCTTACGAGCTTTCATTTTGGAAAGAAATCCCAATATGCAACC 124
 DB 32 AGCGGTGCTAGACCGGGGATCTTTCGACCTGGAATAGAGAGCGCTGGCGGTGCGG 91
 QY 125 TCAATCAAGGCTATGAGAGACTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGAAT 184
 DB 92 TAGCGGACTCTATGCGGGAGAGCTGCGCGCTCGGAGCGGCAACCGGCGGAGCAT 151
 QY 185 CGGTAAATTAATGAGAAAGCAACGCTGAGGCAATTCGGTAGCCGTTATTTTTCGATT 244
 DB 152 CGGTGAGAACCGGTGAGGAGCGCTAATTCGGGCGGCGGCTGCTGTGTGAGT 211
 QY 245 TTTCTCTCTGGGGGTTCTTTGGGCAAGCTGCGGTGCTGCTGATCATGAAGCGATT 304
 DB 212 TCGACTTCTGGGGGCTGATGAGGAGGCGGCGGAGGATCAACCGCGCGCTG 271
 QY 305 ACCGCGCAGAGACTGAACTCCCACTGCTGCTCCCTGCTTCGCGGTGGGCGGCA 364
 DB 272 AGCGGGGCAACCGGAGGCGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 331
 QY 365 TCGAGAAAGACATGAGACTTTTGTATGATGATGATGATGATGATGATGATGATG 424
 DB 332 TCGAAAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 391
 QY 425 ACCGAGGAGGCAATTTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484
 DB 392 ACAACGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
 QY 485 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544

```

Db 452 TCGCGTCGTGGGGCTCGCTGGGGCATCTACCGTCGCCAGCGCGGCGCCCTGATCGGCT 511
Qy 545 TCCGGGTCCTCGCGTGTGAGTTAAACCACTGGGCAATCGGTTCCAGACGGTGTGGAGC 604
Db 512 TTTCTGGGACACAGGGGTCTATGATGTCTCTATGAGGACCCCTTCCATCCGCGCTCAAA 571
Qy 605 AGCGGAGAAATTTGTGAAAATCGTGTGATTTGATGAAATTTGTGTGGCACTCCAAATGCG 664
Db 572 CGCGGAGAAATCTAGGGGGGATGGGATCATGAGAGGGGTGTGCTGACTGGACCGGGTAC 631
Qy 665 GTGACGGGTGGCAAAACCTCAAGTTATTGACCGGTAGAGCAAGCAAGATCGTTT 724
Db 632 GACCATCTGATCGTGTGAGGTGCTCATGAGCGCTCCCAACCGCTTCCGGGAC 691
Qy 725 CTCCAAACAATCTCGCGTGGCA---CTTCCGGTATGAGAGGCGATTCGGGTTCTCGTG 781
Db 692 CGCAGACCCCGCGCCCGTACCCGATGTGCCCATGTGGACTGGGTGTGGCATGCGCC 751
Qy 782 ACCCGCAGAGCGCTGGAATCGGGAGATTATGAAAACGTTGGGGGCGACAGCTGTCAAGC 841
Db 752 GCGCGGACCGCGCGCGCTGAGGCACTACTGCCAGACGGCGCCACCGAGGTGTGT 811
Qy 842 TTTCTGTGCGCGTCTGCGCGCATTTGAGCCCGCGTGTGCGCGTTCCTGTGCGCGCATCG 901
Db 812 TGTCAAGAAACGATCAAGGCGCAAGCG---GCGACACGCTGCTGGCGCTGCGCTTGTG 868
Qy 902 GGGGCGCGCGCGTGTGCTGATTTGGGCGAG-----GATGCGCGCTTCAAGC 946
Db 869 GCGGCAACCAACGCGGTGTCTCGGCGCAGCAAGGCGCATGAGCGCGGGGAGACATCG 928
Qy 947 TTGGGCGCAGAGCTGCGGTTTGGCGCTGTGCGATTTGCTGTGCGCGCGCAAGCTTAAAC 1006
Db 929 TCGGGCCGCGTGGCTTACGCGAAAGCCGACGCGGGATGGCGCTGCGCCGCGATGTGCC 988
Qy 1007 TGGCGATGTGTTCATCTGACACCTCTCGCGCGCGCAATTTGTGCGAGCGGTGAGGAC 1066
Db 989 TGCCCTGTGTGTGATTTGACGCGCGCGGACCGCGTGTGTGCGCGCGCAAGCAAG 1048
Qy 1067 TCGGATGCGCAAGCTCGATTTGGCGGCGACCTTCCAACTTATGAGCGTCCCTCCCGCA 1126
Db 1049 GGGGCTGCGCGCGCAAGTCGCGATTCGCTGGCGGAGCTGTGACGCTGATACCCCGA 1108
Qy 1127 CCGTTTCGATATTTGATGAGGCGTGTGGCGTGTGCGCGCTGCGCATGTGCGCGCG 1186
Db 1109 CCGTGTGATCTCTGTGGGCGAGGCGAGGCGCGCGCGTGTGCGATGTTGCCGCG 1168
Qy 1187 ATCTGTCTAGCGGCGCGCAAAACGCGTGTGTCTGCGATTTGCCACCAAGGCGCTTGTG 1246
Db 1169 ACCGGGTCTGCGCGCATCTCCACGCGCTGTGCGCGCTTGTGCTCCCAAGGAGCGCAAG 1228
Qy 1247 CCATCTCTCTCGCGCAGCAACAGCGCGGGAATCATAGAGCAAGGCGGTGCGAG 1306
Db 1229 CGATGTGTCTCGAGACACTGTCTATGCGCGCGCAACTGCTGCGCCCAAGGCAATCGGT 1288
Qy 1307 CGCAGCACTTTTAAAGCGGCTTATGAGCGGATCGTCCCGCAAAACGAGCACTTTG 1366
Db 1289 CGCGGACACTTGAAGTGGGGATTGTGACACCATGTGTGCGGAGTACCCCGACGCG 1348
Qy 1367 TTGAAGA 1373
Db 1349 CAGACGA 1355

```

RESULT 6

```

US-09-815-242-7800
; Sequence 7800, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

```

```

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EDITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7800
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(873)
; US-09-815-242-7800

Query Match 8.3%; Score 122.8; DB 9; Length 873;
Best Local Similarity 52.8%; Pred. No. 1e-27;
Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

Qy 179 ATGAATCGGTAATTAATGAGAAAGCACCGTGGAGGGCATTCGGTATTTTGT 238
Db 326 AGAGCGGCTGATGCGCATAGCGCGAAGCTGCAAGGATCGCGGTGTGCGCCCT 385
Qy 239 CCGATTTTCTTCTCTCGGCGGTTCTTTTGGGCAAGTGGCGTGGCGATCAAGAG 298
Db 386 TCGAGTTCTCTTCATGAGCGGTTTCGATGAGGCGCATGTCGCGAGCGCTTGTGCG 445
Qy 299 CGATTACCGCGCCACAGACTGAATCTCCACTGCTGTGTCTCCCTGCTTCGGTGTG 358
Db 446 CAGCAACGTGCGCTTGAAGAGCGTGCCTGCTGATCTCTTCTCGCGCTCGGCGG 505
Qy 359 CGCGCATGCAAGAAACATCGAGCTTTTGTATGATGTGTCCATTAACCGCGCTGTC 418
Db 506 CGCGCATGCAAGAAACCGCTATCTCGCTGATGAGATGCGCAAGACCTCGCGGTTCTG 565
Qy 419 AGCGTACCGCGAGCGCATTTTGCCTTCTGTGTATTTTGGCAATCCCATGAGGTG 478
Db 566 CGCGCTGCGGAGAAAGGATCCCGTTCGTCTGATTTAGACGACCGGTTCTACGCG 625
Qy 479 GCGCATGCGCTGTGGGTTTCACTTGGGATCTCACTTTTGGGAACCGCGCGGAGA 538
Db 626 GCGTTTCCGCGAGCTGCGCATGCTGCGCAGGTGATCTGTGCGCAACCGAGCGCTGA 685
Qy 539 TAGGTTCTGCGGCTCTCGCGTGTGAGTTAAACCACTGCGGCAATGCGCTTCCAGACG 598
Db 686 TCGGCTTCCGCGGCTCTCGCGTGTGAGTACGAGACCGTCCGCGAAGCTTCCGGAAG 745
Qy 599 TGCAGCAGCGGAGAAATTTGTGAAAATCGTGTGATTTGATGAAATTTGTGCGCACT 658
Db 746 TCGAGCGTACGAGTTCTCTCTTGAAGATGCGCGCATCGACATGATGTGATGTGCG 805
Qy 659 AATTGCGTCAAGCGGTGCAAA 680
Db 806 AGTTGCGCGCGCGCTGCGCAA 827

```


RESULT 7
US-09-815-242-6828
; Sequence 6828, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 6828
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(867)
US-09-815-242-6828

Query Match 8.2%; Score 121; DB 9; Length 867;
Best Local Similarity 53.0%; Pred. No. 3.7e-27;
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 178 GATGATCGGTAACTGAGAGAGACCGGTGAGGCGATTCCGTAGCCGTATTG 237
Db 328 GATGAGCTGTCTTAACGGGAGAGACCAATGAGAGACAGCTGTTCATTTGGAATC 387
Qy 238 TCCGATTTTCTCTCTGCGCGGTTCTTTGGGACGCTGCGTGGCATATGAAG 297
Db 388 ATGGACGCTAATTTATCATGGGAGATATGGAGACGATTGTGAAAAATCACAGC 447
Qy 298 GCGATTACCGGCGCACAGAGCTGAACCTCCACGCTGCTCCCTGCTCCGCTGCT 357
Db 448 TTGTTAGACGGGCGACAGAAAAAGCATTTACAGTAGATTTTACTGCACTCTGTGCT 507
Qy 358 GCGCGCATGAGAGAGACATGACCTTTTGCATGATGCTGCATTAACCGCGGCTG 417
Db 508 GCCCGTATGCAAGAGAAATTTTTCATTTGATGCAAAATGCGAAATTTGGCGCTTTG 567
Qy 418 CAGGCTCACCGGAGCGCATTTGCCGTTCTGTGTATTTTGGCAATCCACGATGGGT 477
Db 568 CAACGCGATACCAAGACAGGCTTGTCTGTATCTTAAGGATTTGACTGATCAACGACTGGC 527
Qy 478 GCGCGCATGAGCGCTGTGGGCTTCATCTGGGATCTCACTTTTGGCGAATCCGCGCGCAG 537
Db 628 GGTGTACCGCAAGTTTTCGATGATGCGCGATATTTATTTTGGCAGAGCTTCAGAGTTTA 687
Qy 538 ATAGGTTTCTGTGCTCTGCGTGTGAGTAAACCACTGGGCGATGCCCTTCAGACGGT 597

Db 688 ATCGGTTTGTGCGCGCGCTGTATTTGAACAAACGATTCGTCAGAGTTGCCACATGAT 747
Qy 598 GTGCGAGCGCGAGAAATTTGTGTAAGAACTGTGTGATGATGCAATTGTGTCCGCACTC 657
Db 748 TTTCAAAAGCGCGAGTTTCTTTTGAACATGTTTGTGATGACATTTGTCCAAAGAAAT 807
Qy 658 CAATGGCGT 666
Db 808 CTTTGGCGT 816

RESULT 8
US-09-070-927A-458
; Sequence 458, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 982
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 458:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6021 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 458:
US-09-070-927A-458

Query Match 8.2%; Score 121; DB 10; Length 6021;
Best Local Similarity 53.0%; Pred. No. 7.9e-27;
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 178 GATGATCGGTAACTGAGAGAGACCGGTGAGGCGATTCCGTAGCCGTATTG 237
Db 5321 GATGAGCTGTCTTAACGGGAGAGACCAATTTGAGAGACAGCTGTTCATTTGGAATC 5380
Qy 238 TCCGATTTTCTCTCTGCGCGGTTCTTTGGGACGCTGCGTGGCATATGAAG 297
Db 5381 ATGACGCTAATTTATCATGGGAGATATGGAGACGATTTGTGTGAAAAATCACAGC 5440

QY	298	GCGATTACCCGCCCAACAGACTGAAATCCCACTGCTGGTCTCCCGTCCGGAGGT	357
Db	5441	TTGTTTAGAGGGGGAGACAGAAACATTTTACATAGAGATTTTCACTGCATCTGGTGGT	5500
QY	358	GCGCGCATGACGAGGAACAATCGACTTTTGTCTATGATGAGTGTCCATPACCGCGCTGTG	417
Db	5501	GCCCCGTATGCAGAGAGAAATTTTTTCAATGATGCAATGGGAAATTTTGGCGCGCTTGG	5560
QY	418	CAGCGTCACCGCGAGCGCATTTTGCCTTCTGTGATTTTGGCAATCCACGATGGT	477
Db	5561	CAACGCGATACAAAGACAGCGCTTCTGTATCTTTACGATATGACTGATCCACACACTGGC	5620
QY	478	GGCGCCATGGCGCTGTGGGGTTTCAATCGGGGATCTCACTTTGGGAAACCGGCGCGAG	537
Db	5621	GGTGTACCGGAAGTTTTCGATGAGTAGTGCAATATATTTTGGCAGAGCCTCAGAGTTTA	5680
QY	538	ATTAGTTTCTCTGGGTCTCTCGCGTGGTGAATTAACCATGCGCATGCGCTTCCAGACGGT	597
Db	5681	ATCGGTTTTGTGGCGCCGCTGTATTTGAAACAAACGATTCCTCAAGAGATTTCCAGATGAT	5740
QY	598	GTCGACGACGGCGGAGAAATTTGGTGAAAACTGGTGTGATTTGATGGAATTTGTGCGCATTC	657
Db	5741	TTTCAAAAGGCGGAGTTTCTTTTAGAACAATGGTTTGTAGATCAGAGTTGTGCAAGAAAT	5800
QY	658	CAATTGGCT 666	
Db	5801	CTTTTGGCT 5809	

```

RESULT 9
US-09-815-242-7288
; Sequence 7288, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA_011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 7288
;
; LENGTH: 870
;
; TYPE: DNA
;
; ORGANISM: Helicobacter pylori
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (1) .. (870)
;
US-09-815-242-7288

```

Best Local Similarity 50.9%; Pred. No. 1.3e-26;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Oy	93	TTCTTGGAATGAAGAACTCCCAATATGCAACTCCAAATCAAGGCTATGCAAGACCTTGA	152
Db	240	TTTACGGCCCTTAATGATCTCTTAAATTTCTGAGATTAAGAGAGCTATAAACAACGATTTAA	299
Oy	153	GCGGGCTCGAAGCAAGGCCAAATGCGATGAATCGGTATTAATCTGGAGAAGGCACCGTGA	212
Db	300	AAATATGCAAAAAAGGACTAATACGGCCCAAGCTCAGTATCAGCGGTAGGGCTAAATATPA	359
Oy	213	GGGATTCGGGTAGCCCTTATTTTGTCCGATTTTCTCCCTCCGCGGTTCTTTGGGCAC	272
Db	360	CCGCATCCCTTTGCAATCGGTGTTGATTTTAAGCTTTATGGGGGGAGTTTAAAGCTC	419
Oy	273	GGTGCCTCGGTGCGCATCATGAAGGCGATTACCGGCCACAGAGCTGAAATCCCACT	332
Db	420	TGTGAGGGCGAAAGAATCGTAAGAAGCATCAATCGCGCGTCTGCTAAAGAGAAGCGTT	479
Oy	333	GCTGGTCTCCCTGCTTCCGCTGTGCGCGCATGCAAGAAACAATCGAGCTTTGTTCAT	392
Db	480	ATTGATTGTTTCAACGAGTGGGGGGGCTAGAGATCAGAAATCCATTATTCGCTCATGCA	539
Oy	393	GATGCTGCCATTAACCGCGGCTGTGCAAGCCTCACCGCAGGGCGCATTTGACCGTCTCGGT	452
Db	540	AATGGCTAAACGAGCGCGGCTTTGAACCGATTGATGAGGCCAAATCCTCTTTCATTTTC	599
Oy	453	GTATTTGGCATCCCAACGATGGGTGGCGCCATGGCTCGTGGGGTTCAATCGGCATCT	512
Db	600	GCTCTTAAGCATCCCACTTANTGAGGGGGTTAGGCAATCTTTTGTCTTTTATGGGGATCT	659
Oy	513	CACTTTTCGGAAACCCGGCGCGCAGATAGTTTCTGGGTCCTCGCGTGTGAGAGTTAAC	572
Db	660	CATTATCGCAAGCCAGGGCGAGATGATAGCTTTTCGGGGCCCTAAGGATTAAGCAAAAC	719
Oy	573	CACGTGGGATCGCTTCCAGACGGTGTGCACAGAGCGGAGATTTGGTGAACACTGGTGT	632
Db	720	TATAGGGCGGATTTGGCTGAGGGCTTTCAAAACAGCGGATTTTATTTAGACATGGCTT	779
Oy	633	GATTGATGAATTTGG 648	
Db	780	GATTGATATGATTTGTG 795	

```

RESULT 10
US-09-895-913A-97
/ Sequence 97, Application US/09895913A
/ Patent No. US20020160456A1
/ GENERAL INFORMATION:
/ APPLICANT: Kleantous, Harold
/ APPLICANT: Al-Garawi, Amal
/ APPLICANT: Miller, Charles
/ APPLICANT: Tomb, Jean Francois
/ APPLICANT: Oomen, Raymond P.
/ TITLE OF INVENTION: Identification of Polynucleotides
/ TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
/ TITLE OF INVENTION: Genome
/ FILE REFERENCE: 06132/043002
/ CURRENT APPLICATION NUMBER: US/09/895,913A
/ CURRENT FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 08/881,227
/ PRIOR FILING DATE: 1997-06-24
/ NUMBER OF SEQ ID NOS: 368
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 97
/ LENGTH: 1053
/ TYPE: DNA
/ ORGANISM: Helicobacter pylori
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (110)...(976)
/ US-09-895-913A-97

```

Query Match 8.1%; Score 119.2; DB 10; Length 1053;
Best Local Similarity 50.9%; Pred. No. 1.4e-26;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 93 TTCTTGATGAACTCCCAATATGACAACTCAAGGTATGACGACCTTGG 152
DB 349 TTTAGCGCTTAATGATCCCTTTAAATTTCTGATTAAGAGATTAATTAACAACGCAATT 408
QY 153 GCGGGCTGAGCAAGGCAAAATGCGATGAATCGTAATTAATTAATGAGGACCGTGA 212
DB 409 AAAATACGAAAAAGACTTAACCGCAAGCTGATGATGAGGAGGCTTAATTAATCA 468
QY 213 GGGCAATCCGATGAGCGGTAATTTTGTCCATTTTCTTCTGCGGCTTTTGGGCA 272
DB 469 CCGCATGCTTTGACAGATGATGATGATTTTATTTTATTTAGGGGGGAGTTAGGCT 528
QY 273 GGTGCGGTGATGCGGATATGAGGCAATTCACCGGCGACAGAGCTGAATCCCACT 332
DB 529 TGTGAGGGGCAAAAGATGTAAGACAAATCAATCCCGGCTGCTMAAAGAGAGCTT 588
QY 333 GCTGCTCTCCCTGCTTCCGATGATGCGGCAATGAGAAACAATCGAGCTTTGTCA 392
DB 589 ATTTGATTTTTCAGGAGATGAGGAGGCTTGAATGCAAGATCACTTATTCGCTATG 648
QY 393 GATGATGTCATTAACCGGCTGTCAGCGCTCAACGAGGCGCAATTTGCGTCTG 452
DB 649 AATGCTAAACGAGGCGGCTTTGAACGATTAAGTGAAGGCAAACTCCCTTATTT 708
QY 453 GATTTTGGCAATCCCAAGATGAGTGGCCATGCGCTTGTGGGTTATCTGGGCA 512
DB 709 GCTTTAAGCCGATCCCACTTAATGAGGCGCTTGAAGCATTTTGTGTTTGGGATCT 768
QY 513 CACTTTTGGGAAACCGGCGGCAATGATTTTCTGCGGCTTCCGCGGTGGAGTTAC 572
DB 769 CATTTTCAGAGCCAGGCGGCAATGATGAGCTTTGCGGCGCTTGAAGTAAAG 828
QY 573 CACTGGGATGCGCTTCCAGCGATGTCAGAGGCGGAGAAATTTGTAAGAACTG 632
DB 829 TATAGGGGCGATTTGCTGAGGCGCTTCAACAGCGGAAATTTTATTAAGACATG 888
QY 633 GATTGATGAATTTG 648
DB 889 GATTGATGATTTG 904

RESULT 11
US-09-815-242-9854
; Sequence 9854, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9854
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(993)
US-09-815-242-9854

Query Match 7.3%; Score 108; DB 9; Length 993;
Best Local Similarity 49.6%; Pred. No. 4.5e-23;
Matches 276; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

QY 131 AAGGCTATGACAGACCTTGGAGGCGGCTCGAGAGAGCCAAATGCATGATGCTTA 190
DB 344 AAAATATTAAGACAGACTGCGCTCGCGAGAAAGAACGCGAGAAAGCGCGTGG 403
QY 191 TTACTGAGAAAGCACCGTGAAGGCAATTCGCTAGCCGTTATTTTCTTCTTCT 250
DB 404 TGTGATGAAGAGGACGCTTACAGGTATGCGGTTGTGCGCGCGCTTGAATTCGCT 463
QY 251 TCTTGGCGGCTTTTGGGCAAGTGCCTGCTGCTGCTGCTGCTATGAAGGATTCAG 310
DB 464 TCATGGGCGGCTCAATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
QY 311 CCAGAGCTGAATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370
DB 524 GCTGGAAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583
QY 371 AAGACATGAGCTTTTGTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
DB 584 AAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
QY 431 AAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
DB 644 AACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
QY 491 CGTGGGCTTCACTGCGGATTCATTTTGGGAAACCGGCGGCGGCGGCTTCTGCT 550
DB 704 GTTTGCGATCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
QY 551 GTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610
DB 764 GCGCGCGGCTTATGAGCAAAACGCTTGTGAGAGCTTCCGCGAGATTCAGGCGAG 823
QY 611 AGAATTTGTAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
DB 824 AGTTCTGATGAGAAAGCGCTATTTGATGATGATGATGATGATGATGATGATGAT 883
QY 671 CGGTGCAAAACCT 686
DB 884 AGCTGCGAGCATTTCT 899

RESULT 12
US-09-815-242-6133
; Sequence 6133, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23,207,727
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6133
LENGTH: 915
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(915)
US-09-815-242-6133

Query Match
Best Local Similarity 7.0%; Score 102.8; DB 9; Length 915;
Matches 275; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

200 AAGCACCCTGAGGAGGATTCGGGTAGCCGTTATTTTCCGATTTTCTTCTCGGCG 259
335 AAGCACCCTGATGAGATGCGGTTGTCGCTGCGGATTCGAGTTCCTTATGCGCG 394
260 GTTCTTTGGGACCGTCCGTCGCGCATCAGAAAGCGATTACCGCCGACAGAGC 319
395 GTTCAATGGGCTGTTGTTGGGTGACGTTTCGTCGCTGAGAGAGCGCTGAGAG 454
320 TGAACCTCCACCTGCTGCTCCCTGCTTCGCGTGGCGGACATGAGGAAACAATC 379
455 ATAACTGCGCGCTATCTGCTTTCGCTGCGGCGGACATGAGGAAAGACATGA 514
380 GAGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 439
515 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 574
440 TGCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499
575 TGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 634
500 CATCTGCGCATCTCACTTTGCGGAAACCGGCGGACATGATGATGATGATGATGATG 559
635 TGTGCGGCGATCTCAACATCGCTGACCGGAAAGCTTTATCGGCTTTCGCGCTG 694
560 TGTGAGATTAACCACTGGGATGCGCTTCCAGAGGTGTCAGAGGCGGAGAAATTTG 619
695 TTATGCAACAGACCGCTTGGCAAAAACGCGCGCTGATTCAGGCGAGTGAATTCCTG 754
620 TGAACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
755 TCGGAAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 814
680 AAACCTCAAGGATTAATGAGCGGAGAGGCAAGGATGATGATGATGATGATGATG 739
815 GCATTTGCGGAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 874
740 GCGTGGCACTTCGGTGAATGA 761
875 TAGTGTATCCCCCGGTACCGGA 896

RESULT 13
US-09-815-242-9520

Sequence No. US20020061569A1
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9520

LENGTH: 867

TYPE: DNA

ORGANISM: Streptococcus pneumoniae

FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(867)
US-09-815-242-9520

Query Match
Best Local Similarity 6.6%; Score 97.8; DB 9; Length 867;
Matches 249; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

179 ATGAATCGTAATTAATGAGAAAGGACCGTGGAGGCGATTCGGTATGCTTATTTGT 238
332 ATGAAGCCGTTTGTACAGAAATGCTCTTATTAAGGTGACATGCTGTGGATTA 391
239 CCGATTTTCTTCTTCTGCGGCTTCTTTGGGACGATCGGCTGGTCCGATCAATGAAG 298
392 TGATTTTAATCTTATCATGATGATGATGATGATGATGATGATGATGATGATGATG 451
299 CGATTCACCGGCGACAGAGTGAATCCCACTGCTGCTGCTTCCCTGCTTCCGCTG 358
452 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 511
359 CCGGATGAGGAAAGCAATGAGCTTTGATGATGATGATGATGATGATGATGATGATG 418
512 CCGATGAGGAAAGCAATGAGCTTTGATGATGATGATGATGATGATGATGATGATG 571
419 AGCTACCGGAGGCGCATTTTCCCTGCTGATGATGATGATGATGATGATGATGATG 478
572 AAGCTATTAAGTGTGCTCTTTTACCTGATGATGATGATGATGATGATGATGATG 631
479 GCGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538
632 GTGTGACAGCTTTTGTGATGAGAAAGCAATGATGATGATGATGATGATGATGATG 691

Qy 539 TAGTTTCTGGGTCCTGGCGTGTGAGCTTAACCACTGGGCGATGCGCTTCAGACGGTG 598
Db 632 TTGGTTTGTCTGGGGCGTGTGATTTGAATAACGGTGTGTAAGCTTGCGTAGAGATT 751
Qy 599 TGCACAGCGGAGAAATTTGGTGAATACTGGGTGTGATTTGATGAAATTTGTGTGCCACTCC 658
Db 752 TCCAAAGGCAAAATTCCTATTAGAACATGAGCTTTGTGATGCTATTGTCAAAAGAGAG 811
Qy 659 AATTGCGTCAAGCGGAGCA 679
Db 812 ACTTACGATACGATTCGTA 832

RESULT 14
US-09-974-300-5531
Sequence 5531, Application US/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:
APPLICANT: Beirka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085-500-US
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5531
LENGTH: 654
TYPE: DNA
ORGANISM: Bacillus clausii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(654)
OTHER INFORMATION: n = A,T,C or G
US-09-974-300-5531

Query Match 6.6%; Score 96.8; DB 10; Length 654;
Best Local Similarity 51.8%; Pred. No. 1.2e-19;
Matches 288; Conservative 0; Mismatches 264; Indels 4; Gaps 3;
Qy 129 TCAAGGCTATGACAGACCTTGGAGCGGCTTGAAGCAAGGCCAATGCGATGATCGGT 188
Db 84 TCCAGTTATAGAGAAAGCTTGAGAGCGACCGTAAACCCGTTGAAATGAAGCGGT 143
Qy 189 AATTACTGAGAGGACCGTGGAGGCAATCCGGTAGCCGTT-ATTTTGTCCGATTTT 247
Db 144 GGTCACTGTGAAGAAAGATTAATGATATCCGTTGATTCATTCAGATTTATGAGCTCA 203
Qy 248 CTTTCTCGGCGGCTTTTGGGCAAGTCCGCTCGGTGCGATCATGAAGCGATTCACC 307
Db 204 ATTTTCGATGGCGAGCATGGGCTCGGGGTTGGGAAAGCTAACGAGGGCAATTGAT 263
Qy 308 GCGCCACAGAGCTGAATCTCCACTGCTGG--TCTCCCTGCTTCGGGTGTGCGCGAT 365
Db 264 TGGGCACCTGAGACGAGCGCTTCGCTTNAATTTTGTGCTGCTCAGGGGAGCGCGCAT 323
Qy 366 GCAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424
Db 324 GCAAGAGGATGCTCAGACCTTATGCAATGCGAAGGAGGCGCGCTTGAAGGCTC 383
Qy 425 ACCGAGAGCGCATTTGCGTCTGATGATTTGCGCAATCCACGATGGGTGGCGCA 484
Db 384 TGAACCGAGCAGAGGCGCTTATCATTTGCGTTATGACCAACCGAGCGAGCGCGGT 443
Qy 485 TGGCTCTGTGGGTTCACTGTGGGATCTCACTTTTGGGAAACCGGCGCGAGATAGTT 544
Db 444 CCGCTAGTTTGGCGCTCTCTTGGCGATTAATTTTGGCGAGCAAAAGCACTATTTGCT 503

Qy 545 TCCTGGTCTTCGCGTGTGAGCTTAACCACTGGGCGATGCGCTTCAGACGGTGTGACG 604
Db 504 TGCCTGGAGGCGCATTTTGAACAGACCATTCGTCAGGAATTCCTGTAAGCTTCAGA 563
Qy 605 AGCGGAGAAATTTGGTGAATACTGGGTGTGATTTGATGAAATTTGTGTGCCACTCCAATTGC 664
Db 564 CAGCGAAATTTCTCTTGAGAGATGGCCAGCTTGAATGATGATTCATTCAGAGCTAAGAAATGA 623
Qy 665 GTTCAGCGGTGGCA 680
Db 624 AAGAAACGCTGACTGA 639

RESULT 15
US-09-815-242-7115
Sequence 7115, Application US/09815242
Patent No. US2002061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7115
LENGTH: 891
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(891)
US-09-815-242-7115

Query Match 6.5%; Score 96.2; DB 9; Length 891;
Best Local Similarity 49.5%; Pred. No. 2.1e-19;
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

Qy 179 ATGAATCGTAACTACTGAGAGCAACCGTGGAGGCGATTCGGGTAGCGCTTATTTGT 238
Db 320 AAGATGCGCTAATTAATGACAGATACCTTTAATATGCAATGCTTGTGGCTGAT 379
Qy 239 CCGATTTTCTCTCGCGGCTTTTGGGCAAGTCCGCTCGGTGCGATCATGAAG 298
Db 380 CGAATTTGCTTTATGGGCGGTTCAATGGGTCTGTAGTTGGTGAATAATTTGTTAAG 439
Qy 299 CGATTACCGGCGACAGAGTGAATCTCCACTGCTGTGTCTCCCTGCTTCGGTGTG 358
Db 440 CGGCTGAAGAGCGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
Qy 359 GCGCATGAGAGAGCAATGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGAT 418

Db	500	CTCGTATGCAGAGACATTATTCTTCTTAATGCAAAATGCAAAAATAGTGCCTGACTTG	559
Qy	419	AGCGTCAACCGGAGAGCCGATTTGCGCTCCGCTGTATTTGGCAATCCCAAGATGGTG	478
Db	560	CTCAAAATGCGTGAAGGGGTGTGCAATTTTTCAGATTAACGGATCGACTTTAGGCG	619
Qy	479	GCGCCATGCGCTCGTGGGGTTCATCTGGGCATCTCACTTTTGGGAACCGGCGGCAGA	538
Db	620	GCGTATCAGCCAGTTTTCGATGTTAGGGGATTTAATATTCGAGCCAAAAGCCTTA	679
Qy	539	TAGGTTCTGCGTCTCGCTGTGAGTTAACCATGCGCATGCGCTTCCAGACGGTG	598
Db	680	TTGGTTTTCAGGGCCACGCTAATTAATGAACAACCTGCGTGAATAAATTGCCAGAGGTT	739
Qy	599	TGCAGCAGGCGGAGATTGTGTAATACTGTGTGATGATGGAATTTGTGCGCCTCC	658
Db	740	TCCAACGTAGTAGTTTCTACTTGAGAAAGGGCAATTGATATGATCGTGAACGTTTCAG	799
Qy	659	AATTGCGTGCAGCGTGSCAA	679
Db	800	AAATCGTCAAACTTTAGCAA	820

Search completed: November 13, 2003, 06:16:54
 Job time : 912.783 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 00:15:23 ; Search time 75 Seconds

(without alignments)
1039.130 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKPEPTWVGMEHTSALT.....NNPERAGRDSRFRPERLQAQ 491

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_13Jun03.*
1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2448	100.0	491	22	AA045789
2	2445	99.9	491	22	AA090675
3	2411	98.5	484	22	AA076522
4	1047	42.8	495	22	AA081130
5	400	16.3	288	22	AA035332
6	384	15.7	289	19	AA098434
7	384	15.7	289	22	AA035792
8	383	15.6	225	23	AA051006
9	375	15.3	330	22	AA038359

10	374	15.3	304	22	AA034637	E. coli cellular p
11	368	15.0	288	22	AB054092	Lactococcus lactis
12	368	15.0	296	22	AA035619	Haemophilus influenzae
13	365	14.9	291	23	AB028019	Streptococcus poly
14	364	14.9	312	20	AA036893	Protein involved i
15	362	14.8	279	23	AB030977	Helicobacter pylori
16	360.5	14.7	1106	22	AB025523	Novel human diagno
17	354.5	14.5	285	22	AA083049	S. epidermidis ope
18	354.5	14.5	293	23	AB038958	Staphylococcus epi
19	344.5	14.1	292	24	AB076030	N. gonorrhoeae ami
20	343.5	14.0	285	22	AA037025	Staphylococcus aur
21	343.5	14.0	285	22	AA073368	Staphylococcus aur
22	340.5	13.9	282	22	AA033928	Staphylococcus aur
23	335	13.7	288	23	AB028020	Streptococcus poly
24	334	13.6	288	22	AA037772	Streptococcus pneu
25	334	13.6	288	22	AA036025	Streptococcus pneu
26	334	13.6	288	22	AA010064	CPE 67 protein seq
27	334	13.6	288	24	AA000780	S. pneumoniae type
28	329.5	13.5	290	22	AA036305	Pseudomonas aerugi
29	319.5	13.1	255	20	AA034658	Chlamydia pneumoniae
30	318.5	13.0	294	23	AB048048	Listeria monocytog
31	307	12.5	527	22	AB046609	Putative P. abyss
32	300.5	12.3	318	23	AB046047	Listeria monocytog
33	291.5	11.9	256	23	AB028022	Streptococcus poly
34	291.5	11.9	280	22	AA083050	S. epidermidis ope
35	289	11.8	317	18	AA027453	Magnetospirillum s
36	288	11.8	315	22	AA035421	Haemophilus influe
37	288	11.8	315	24	AA030458	Haemophilus influe
38	285.5	11.7	540	23	AB061186	Blidobacterium lo
39	285	11.6	527	23	AB096725	S. coelicolor AcCB
40	283.5	11.6	315	23	AB038941	Staphylococcus epi
41	282	11.5	875	19	AA048309	Plasm sativum AcCa
42	277	11.3	309	22	AA034396	Staphylococcus aur
43	275	11.2	314	22	AA037315	Staphylococcus aur
44	275	11.2	526	24	AB057631	S. mutaymaensis A
45	274	11.2	314	22	AA072365	Staphylococcus aur

ALIGNMENTS

RESULT 1	AA045789	standard; Protein; 491 AA.
ID	AA045789	
AC	AA045789	
XX		
DT	15-MAR-2001	(first entry)
XX		
DE	C. glutamicum accDA protein.	
XX		
KW	L-amino acid; accDA; coryneform microorganism; L-lysine; animal feed; medicine; pharmaceutical industry.	
XX		
OS	Corynebacterium glutamicum.	
XX		
PN	EPI055725-A2.	
XX		
PD	29-NOV-2000.	
XX		
PF	10-MAY-2000; 2000EP-0109842.	
XX		
PR	27-MAY-1999; 99DE-1024365.	
XX		
PA	(DEGS) DEGUSSA-HUELS AG.	
XX	(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.	
PI	Tilg Y, Eggeling L, Birkmann B, Sahm H, Moeckel B,	
DR	WPI; 2001-042411/06.	
XX	N-PSDB; AAC82732; AAC82733.	
PT	Cloned Corynebacterium glutamicum accDA gene useful for producing	

transformed coryneform bacteria producing increased yields of L-amino acids, especially L-lysine

Claim 3; Page 15-17; 20pp; German.

This invention describes a novel cloned Corynebacterium glutamicum DNA (1), replicable in coryneform microorganisms coding for an accD gene. The invention also describes (1) a polypeptide derived from (1) having a fully defined 491 aa sequence; (2) coryneform microorganisms transformed with one or more copies of (1); (3) the shuttle vector pZlacCDa contained in Corynebacterium glutamicum DSM 12785; and (4) a process for producing L-amino acids, comprising culturing a coryneform bacterium that overexpresses the accD gene. Coryneform bacteria transformed with (1) so that they overexpress the accD gene are useful for producing L-amino acids, especially L-lysine, which are useful in animal feeds, in human medicine and in the pharmaceutical industry.

Sequence 491 AA:

Query Match 100.0%; Score 2448; DB 22; Length 491;

Best Local Similarity 100.0%; Pred. No. 7.66-215; Mismatches 0; Indels 0; Gaps 0;

Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VEKRPPTVMWGMHETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60
1 VEKRPPTVMWGMHETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60

61 ESVITGEGTVGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120
61 ESVITGEGTVGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120

121 RMOEDNRAFWMVVSTTAAVQHRRAHLPVLVYLNPTMGAMASWSSGHLTFEPGQOI 180
121 RMOEDNRAFWMVVSTTAAVQHRRAHLPVLVYLNPTMGAMASWSSGHLTFEPGQOI 180

181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGVSPLOQAATAVAKTLKVIQVEATDR 240
181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGVSPLOQAATAVAKTLKVIQVEATDR 240

241 FSPPTPGVALPVMEAIARSRDPQPGIGIEMETLGADVVKISGARAGALSPAVRALARI 300
241 FSPPTPGVALPVMEAIARSRDPQPGIGIEMETLGADVVKISGARAGALSPAVRALARI 300

301 GGRPVVLTIGDRRFTLGPQELRPARRGISLARELNLPVSIITDTSAGAEISQAABELGIAS 360
301 GGRPVVLTIGDRRFTLGPQELRPARRGISLARELNLPVSIITDTSAGAEISQAABELGIAS 360

361 STARTLSKLIDAPLPVSVITIGGVGGGALAMLPAADVVAENAMLSALPPEGASAILFR 420
361 STARTLSKLIDAPLPVSVITIGGVGGGALAMLPAADVVAENAMLSALPPEGASAILFR 420

421 DTNHAABEIERQGVAAHALISQGLIDGIVAEETHEFVEEILGTISNALSELNNPERAGRD 480
421 DTNHAABEIERQGVAAHALISQGLIDGIVAEETHEFVEEILGTISNALSELNNPERAGRD 480

481 SRFTFRERLAQ 491
481 SRFTFRERLAQ 491

RESULT 2

AAAG90675 ID AAG90675 standard; Protein; 491 AA.

AAAG90675;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 4429.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH65894.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 4429; 246pp + Sequence listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 491 AA:

Query Match 99.9%; Score 2445; DB 22; Length 491;

Best Local Similarity 99.8%; Pred. No. 1.46-214; Mismatches 0; Indels 0; Gaps 0;

Matches 490; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 VEKRPPTVMWGMHETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60
1 MEKRPPTVMWGMHETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60

61 ESVITGEGTVGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120
61 ESVITGEGTVGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120

121 RMOEDNRAFWMVVSTTAAVQHRRAHLPVLVYLNPTMGAMASWSSGHLTFEPGQOI 180
121 RMOEDNRAFWMVVSTTAAVQHRRAHLPVLVYLNPTMGAMASWSSGHLTFEPGQOI 180

181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGVSPLOQAATAVAKTLKVIQVEATDR 240
181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGVSPLOQAATAVAKTLKVIQVEATDR 240

241 FSPPTPGVALPVMEAIARSRDPQPGIGIEMETLGADVVKISGARAGALSPAVRALARI 300
241 FSPPTPGVALPVMEAIARSRDPQPGIGIEMETLGADVVKISGARAGALSPAVRALARI 300

301 GGRPVVLTIGDRRFTLGPQELRPARRGISLARELNLPVSIITDTSAGAEISQAABELGIAS 360
301 GGRPVVLTIGDRRFTLGPQELRPARRGISLARELNLPVSIITDTSAGAEISQAABELGIAS 360

361 STARTLSKLIDAPLPVSVITIGGVGGGALAMLPAADVVAENAMLSALPPEGASAILFR 420
361 STARTLSKLIDAPLPVSVITIGGVGGGALAMLPAADVVAENAMLSALPPEGASAILFR 420

DB 361 STARTLSKIDAPLPTVSIIIGGVGGALAMLPADVVAANAMLSALPFGASAILPR 420
 QY 421 DNNHAEIIRERGQVQAHALLSGGLIDGIVAEFEHFEVEILGTISNALSELNNPERAGRD 480
 DB 421 DNNHAEIIRERGQVQAHALLSGGLIDGIVAEFEHFEVEILGTISNALSELNNPERAGRD 480
 QY 481 SRFTRFERLAQ 491
 DB 481 SRFTRFERLAQ 491

RESULT 3
 AAB76522
 ID AAB76522 standard; Protein; 484 AA.
 XX
 AC AAB76522;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein SHQ ID NO:26.
 XX
 KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering.
 XX
 OS Corynebacterium glutamicum.
 XX
 PV MO200100805-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000MO-IB00926.
 XX
 XX 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031563.
 PR 09-JUL-1999; 99DE-1032122.
 PR 09-JUL-1999; 99DE-1032124.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032182.
 PR 09-JUL-1999; 99DE-1032190.
 PR 09-JUL-1999; 99DE-1032191.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032212.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 14-JUL-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 XX (BADI) BASF AG.

XX
 PI Pompejus M, Kroege B, Schroeder H, Zelder O, Haberman G;
 XX
 DR MPI: 2001-071486/08.
 DR N-PSDB; AAF67755.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation -
 XX
 PS Claim 20; Page 175-177; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 484 AA;
 XX

Query Match 98.5%; Score 2411; DB 22; Length 484;
 Best Local Similarity 100.0%; Pred. No. 1.0e-211;
 Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 MYMGHEHTSALTLTDSVLPDPSFISWNETPOYDNLNOGYAEFTLERARSKACDESVINGE 67
 DB 1 MYMGHEHTSALTLTDSVLPDPSFISWNETPOYDNLNOGYAEFTLERARSKACDESVINGE 60
 QY 68 GTVEGIPVAVILSDPSFLGSGIGTVAASVIRKAIHRATELKLPLVSPASGARGMOEDNR 127
 DB 61 GTVEGIPVAVILSDPSFLGSGIGTVAASVIRKAIHRATELKLPLVSPASGARGMOEDNR 120
 QY 128 AFVWVWSTTAIVQRRRAHLPLVLYLRNPTMGAMASVSGSGHUTFAEPGAQIGLGRV 187
 DB 121 AFVWVWSTTAIVQRRRAHLPLVLYLRNPTMGAMASVSGSGHUTFAEPGAQIGLGRV 180
 QY 188 VELTTGHALPDGVQOAEHLVKTGVYIDGIVSPLOJRAAFAKTLKYQPEATDRFSPPTPG 247
 DB 181 VELTTGHALPDGVQOAEHLVKTGVYIDGIVSPLOJRAAFAKTLKYQPEATDRFSPPTPG 240
 QY 248 VALPMEAIARSRDPOREGEIMETLGADYVKLSGARAGALSAPVRVALARIGRPPVL 307
 DB 241 VALPMEAIARSRDPOREGEIMETLGADYVKLSGARAGALSAPVRVALARIGRPPVL 300
 QY 308 IGDRRFTLPQOELRPARGISLARELMLPIVSIITDTSGAELSQAAEELGIASSIARTLS 367
 DB 301 IGDRRFTLPQOELRPARGISLARELMLPIVSIITDTSGAELSQAAEELGIASSIARTLS 360
 QY 368 KLIIAPLPTVSIIIGGVGGGALAMLPADLYAAENMLSLPPEGASAIIFPDNNHAE 427
 DB 361 KLIIAPLPTVSIIIGGVGGGALAMLPADLYAAENMLSLPPEGASAIIFPDNNHAE 420
 QY 428 IIERGQVQAHALLSGGLIDGIVAEFEHFEVEILGTISNALSELNNPERAGRDSTRFPE 487
 DB 421 IIERGQVQAHALLSGGLIDGIVAEFEHFEVEILGTISNALSELNNPERAGRDSTRFPE 480
 QY 488 RLAAQ 491
 DB 481 RLAAQ 484

RESULT 4
 AAG81130
 ID AAG81130 standard; Protein; 495 AA.
 XX
 AC AAG81130;
 XX

[illegible]

Db	306	GGSTVGPAMLRREARRGMALAEELCTPLVLVIDAAGPALISAAEQQGLAGQIAHCAELVLT	365
QY	372	APLFTVASITTIQGGGGALAMLPAIDLVYAAENAMLSALPEPGASAILFRDTNHAETIER	431
Db	366	LDITVTSITLLQGGGGPALAMLPA DRVIALAHGMLAPLPEPGASAIYVRDTNHAELAA	425
QY	432	QGVQAHALLSGGLIDGIVATEHEFVEETLG--TISNALSELDN----NPERAGRDSR	482
Db	426	QGIRSAIDLTKSGIIVDTIIVEYPDADAEPIEFALRLSNIAAEVHALRKIPAPERLAA--TR	483
QY	483	FTPRERL 489	
Db	484	LQRYRRI 490	
RESULT 5			
AAU35332			
ID	AAU35332	standard; Protein, 288 AA.	
XX			
AC	AAU35332;		
XX			
DT	14-FEB-2002	(first entry)	
XX			
DE	Enterococcus faecalis cellular proliferation protein #619.		
XX			
KW	Antisense; prokaryotic cellular proliferation protein;		
XX	antibiotic; antibacterial; drug design.		
OS	Enterococcus faecalis.		
XX			
PN	WO200170955-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	21-MAR-2001; 2001WO-US09180.		
XX			
PR	21-MAR-2000; 2000US-191078P.		
XX			
PR	23-MAY-2000; 2000US-206848P.		
XX			
PR	26-MAY-2000; 2000US-207727P.		
XX			
PR	23-OCT-2000; 2000US-242578P.		
XX			
PR	27-NOV-2000; 2000US-253625P.		
XX			
PR	22-DEC-2000; 2000US-257931P.		
XX			
PR	16-FEB-2001; 2001US-269308P.		
XX			
PA	(ELIT-) ELITRA PHARM INC.		
XX			
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;		
XX	Yamamoto RT, Xu HH;		
XX			
DR	WPI; 2001-611495/70.		
XX			
DR	N-PSDB; AAS53191.		
XX			
PT	New polynucleotides for the identification and development of		
XX	antibiotics, comprise sequences of antisense nucleic acids -		
XX			
PS	Example 3; Seq ID No 10925; 511pp; English.		
XX			
CC	The invention relates to antisense inhibitors of genes essential to		
CC	prokaryotic cellular proliferation, their use in identifying the		
CC	genes, their use in the discovery of novel antibiotics, the essential		
CC	genes themselves and the encoded proteins. The prokaryotes used are		
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella		
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The		
CC	invention is also useful for the identification of potential new targets		
CC	for antibiotic development. The antisense nucleic acids can also be used		
CC	to identify proteins used in proliferation, to express these proteins,		
CC	and to obtain antibodies capable of binding to the expressed proteins.		
CC	The proteins can be used to screen compounds in rational drug discovery		
CC	programmes. The antisense nucleic acid sequence is also useful to screen		
CC	for homologous nucleic acids which are required for cell proliferation in		
CC	a wide variety of organisms. The present sequence represents an		
CC	essential prokaryotic cellular proliferation protein.		

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 288 AA;

Query Match 16.3%; Score 400; DB 22; Length 288;
Best Local Similarity 42.6%; Pred. No. 5.2e-28;
Matches 89; Conservative 39; Mismatches 79; Indels 2; Gaps 2;

QY 25 LDPOSFISM-NETPOYDNLN-OGVAETLERARSKAKCDESVITGEGTGVAVILSDP 82
DB 73 VDEKSFHMDSELYTKDPLNFPGYLEKIEKMEKGTGLDEAVLTGEGATTEGCAVALGINDA 132
QY 83 SFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTIAVQHRH 142
DB 133 NFIMSGMGTIVVEKTRFEEATEKGLPVIFVTASGARMQEGISLMQMAKISALQHR 192
QY 143 REAHLPFLVLRNPTMGAMASWGSSGHLTPAEPGAOIGFLGPRVVELTTGHALPDGVOQ 202
DB 193 NKAGLLVLTULTDPTTGVTASFAMDGDIILAEPSGLIGFAGRVIEOTIRQELPDPFOK 252
QY 203 AENLVKGVIDGIVSPLOLRAVAATLKV 231
DB 253 AEFLEHGFVDPQIVPRNLLRQLRSLDLRL 281

RESULT 6
AAW98434
ID AAW98434 standard; Protein; 289 AA.

XX AAW98434;

XX 31-MAR-1999 (first entry)

XX H. pylori GHPD 451 protein.

XX GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleanthouse H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14153.

XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases

XX Claim 8; Page 738-739; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPD protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for

CC detection and diagnosis.

XX Sequence 289 AA;

Query Match 15.7%; Score 384; DB 19; Length 289;
Best Local Similarity 39.3%; Pred. No. 1.5e-26;
Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;

QY 26 DPDSFISNENPQYDNLNQGVAETLERARSKAKCDESVITGEGTGVAVILSDPFL 85
DB 85 DPLNFVD-----KESYKORIKKYEKTNPPSSVISEAKINRMPLOIVDFPFSFM 134
QY 86 GSGISGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTIAVQHRH 145
DB 135 GSGISGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTIAVQHRH 194
QY 146 HLPFLVLRNPTMGAMASWGSSGHLTPAEPGAOIGFLGPRVVELTTGHALPDGVOQ 205
DB 195 KLPFLISLSDPTTGVTASFAFLDGLIIEFGAMIGFAGRVIEOTIRQELPDPFOK 254
QY 206 LVKTGVIDGIVSPLOLRAVA-----KTLKV 231
DB 255 LIEHGLIDMIVHRDLKKTLSDLIAMMTHKTSKI 288

RESULT 7
AAU35792
ID AAU35792 standard; Protein; 289 AA.

XX AAU35792;

XX 14-FEB-2002 (first entry)

XX Helicobacter pylori cellular proliferation protein #105.

XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.

XX Helicobacter pylori.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS53651.

XX Example 3; Seq ID No 11385; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

XX FR2807446-A1.
 XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-0004630.
 XX 11-APR-2000; 2000FR-0004630.
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification of Lactococcus
 XX lactis and related species -
 XX Claim 6; SEQ ID No 794; 2504bp; French.
 XX The present invention is related to a Lactococcus lactis nucleotide
 XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 XX nucleic acid sequence is useful in the detection and/or amplification of
 XX nucleic acid sequence, particularly to identify Lactococcus lactis or
 XX related species. The proteins of the invention are useful for the
 XX biosynthesis or biodegradation of a composition of interest. The
 XX invention helps research in lactic bacteria, particularly useful in the
 XX production of yogurt and cheese.
 XX Note: The sequence data for this patent is based on equivalent patent
 XX WO200177334 (published 18-OCT-2001) which is available in electronic
 XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 288 AA;

Query Match 15.0%; Score 368; DB 23; Length 288;
 Best Local Similarity 40.0%; Pred. No. 4.4e-25;
 Matches 86; Conservative 37; Mismatches 90; Indels 2; Gaps 2;

QY 24 VLDPSFISMNETPOVDN-LN-QGYAETLERARSKACDESVITGEGTVEGIPVAVILSD 81
 DB 72 VADKDSFVEMFTGIESKNPLDPGGPEKLAATKARTGDEAVITGATTKGQKTLAAMD 131
 QY 82 FFLGSLGTGVSVRIMKAIHRAETELKPLVSPASGARGMOEDNRAFMVMTSITAAVQR 141
 DB 132 STFIWASMGTVGKXKTRLFYVATTEKLPITVFTASGARGMOEGIMSLMQAKTSAAYKR 191
 QY 142 HREAHLPFLVYLNPPTMGAMASWSSGHLTPAEPQAQIGFLGPRVVELTTGHALPDGVQ 201
 DB 192 HSNAGLFYITVLTPTTGVTASFASLGDIIILAEPSLIGFAGRRVIEQTVQTLPPDFQ 251
 QY 202 QAEINVKTGVIDGIVSPDLQRAAVAKTLKVIQVPE 236
 DB 252 KAEFLINHGFDVAIVKRTLRKMLALLETHTVE 286

RESULT 12
 AAU35619
 ID AAU35619 standard; Protein; 296 AA.
 XX AAU35619;
 XX 14-FEB-2002 (first entry)
 XX Haemophilus influenzae cellular proliferation protein #260.
 XX Antisense; prokaryotic cellular proliferation protein;
 XX antibiotic; antibacterial; drug design.
 XX Haemophilus influenzae.
 XX WO200170955-A2.
 XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 XX 23-MAY-2000; 2000US-206848P.
 XX 26-MAY-2000; 2000US-207727P.
 XX 23-OCT-2000; 2000US-245278P.
 XX 27-NOV-2000; 2000US-253625P.
 XX 22-DEC-2000; 2000US-257931P.
 XX 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Haeselbeck R, Ohlsen KU, Zysek JD, Wall D, Trawick JD, Carr GJ;
 XX Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 XX N-PSDB; AAS53478.
 XX New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 11212; 511bp; English.

The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

Sequence 296 AA;

Query Match 15.0%; Score 368; DB 22; Length 296;
 Best Local Similarity 36.8%; Pred. No. 4.6e-25;
 Matches 81; Conservative 45; Mismatches 80; Indels 14; Gaps 3;

QY 25 LDPSFISMNETPOY-DNLQGYAETLERARSKACDESVITGEGTVEGIPVAVILSDFS 83
 DB 79 LEFDIILKFKOLKXKRYKORINAAQKRETEK-----DALITWGTLYNNEPIVVAASNFA 130
 QY 84 FLGSLGTGVSVRIMKAIHRAETELKPLVSPASGARGMOEDNRAFMVMTSITAAVQRH 143
 DB 131 FMGSSMGSVGAKVKAIAKRAMENCPVCPASGARGMOEDNRAFMVMTSITAAVQRH 190
 QY 144 EAHLPFLVYLNPPTMGAMASWSSGHLTPAEPQAQIGFLGPRVVELTTGHALPDGVQA 203
 DB 191 EKGVPFISVLTPTTGVTASFASLGDIIILAEPSLIGFAGRRVIEQTVREKLPFGQRS 250
 QY 204 ENLVKTGVIDGIVSPDLQRAAVAKTLKVIQVPEATDFSP 243
 DB 251 EFLLEKGAIDMIVKSEMRQTLASVLS-----KLTPQSP 285

RESULT 13
 ABP28019
 ID ABP28019 standard; Protein; 291 AA.
 XX ABP28019;
 XX

```

DT      02-JUL-2002 (first entry)
XX
DE      Streptococcus polypeptide SEQ ID NO 5214.
XX
KW      Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW      group A streptococcus; Streptococcus pyogenes; antibacterial;
XX      antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS      Streptococcus agalactiae.
XX
PN      WO200234771-A2.
XX
PD      02-MAY-2002.
XX
PF      29-OCT-2001; 2001WO-GB04789.
XX
PR      27-OCT-2000; 2000GB-002633.
XX      24-NOV-2000; 2000GB-002827.
PR      07-MAR-2001; 2001GB-0005640.
XX
PA      (CHIR-) CHIRON SPA.
XX      (GENO-) INST GENOMIC RES.
PI      Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
PI      Tettelein H;
DR      WPI: 2002-352536/38.
DR      N-PSTDB; ABN68650.
XX
PT      New Streptococcus protein for the treatment or prevention of infection
PT      or disease caused by Streptococcus bacteria, such as meningitis, and
PT      for detecting a compound that binds to the protein -
XX
XX      Claim 1; Page 3685; 4525pp; English.
XX
XX      The invention relates to a protein (ABP25413-ABP30895) from group B
CC      streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC      (Streptococcus pyogenes), comprising one of 5483 sequences (51), given in
CC      the specification. The proteins have antibacterial and anti-inflammatory
CC      activity. (1), nucleic acid encoding (1), ABN6044-ABN71526 and
CC      antibodies that bind (1) are used in the manufacture of medicaments for
CC      the treatment or prevention of infection or disease caused by
CC      Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC      Nucleic acids encoding (1) are used to detect Streptococcus in a
CC      biological sample. (1) is used to determine whether a compound binds to
CC      (1). A composition comprising (1) or a nucleic acid encoding (1), may be
CC      used as a vaccine or diagnostic composition. The disease caused by
CC      Streptococcus that is prevented or treated may be meningitis. Nucleic
CC      acid encoding (1) may be used to recombinantly produce (1) and may be
CC      used in gene therapy. Antibodies to (1) are used for affinity
CC      chromatography, immunoassays, and distinguishing/identifying
CC      Streptococcus proteins.
XX
SQ      Sequence      291 AA;

Query Match          14.9%; Score 365; DB 23; Length 291;
Best Local Similarity 39.2%; Pred. No. 8.4e-25;
Matches 82; Conservative 38; Mismatches 83; Indels 6; Gaps 3

   25 LDPDSF---ISNNEFPQYDNLN-QGYAETLEARSKAKKDESVITGEGIVEGI PVAVLIS 80
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    75 VDSESFEEELFGIEIK--DPLNFPPYREKLATROKTMDENVTVGLARIKQTALAIM 132
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :

   81 DFSFLGSIGTAVSRIMKAIRATELKLPLLVSPASGARQENRFAVMVVMTAAVO 140
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    133 DSHFTMAMSGIVGEKRLTRFLAELEKLPITYIFPASGARHQEGTMSLMQAOKYSAAVK 192
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :

   QY  141 RHREAHLFLLVLRNPWTMGAMASWGSSGHLPFAEPGAQIGFLGRVVELITGHALPDGV 200
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db    193 RHSNOGLFYLLITLTPPTGGVTFASAMEGDIIILAEPOALVGPAGRVLIETTYREDLPBGF 252
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :

QY  201 QOAENVTKTGVIDGVSPLOLRAPAAVKTL 229
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | :

```

ID	Accession	Standard	Protein	Length
253	QKAEFLHGFVDALINRTELKDCAIQLI	261		
RESULT 14				
ID	AA036893	standard	Protein	312 AA.
XX	AA036893			
AC	AA036893			
XX				
DT	07-OCT-1999	(first entry)		
XX				
DE	Protein involved in intermediate metabolism of fatty acids.			
XX				
KW	Vaccine; eye disease; conventional trachoma; nongonococcal trachoma; paratrachoma; inclusion conjunctivitis; genital disease; periorbital; nongonococcal urethritis; epididymitis; cervicitis; salpingitis; bathrothelitis; pneumopathy; venereal lymphogranulomatosis.			
XX				
OS	Chlamydia trachomatis.			
XX				
PN	W09928475-A2.			
XX				
PD	10-JUN-1999.			
XX				
PF	27-NOV-1998;	98WO-IB01939.		
XX				
PR	04-NOV-1998;	98US-0107077.		
XX				
PR	28-NOV-1997;	97FR-0015041.		
XX				
PR	17-DEC-1997;	97FR-0016034.		
XX				
PA	(GENSET) GENSET.			
XX				
PI	Griffais R;			
XX				
DR	WPI; 1999-371125/31.			
XX				
XX	Genome sequence of Chlamydia trachomatis			
XX	Disclosure; Page 761; 1755pp; English.			
XX	AA036754-37949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AA021425). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nongonococcal trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, periorbital, bathrothelitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.			
XX	Sequence	312 AA;		
XX	Query Match	14.9%;	Score 364;	DB 20; Length 312;
XX	Best Local Similarity	39.3%;	Pred. No. 1.2e-24;	
XX	Matches	81;	Conservative % 35;	Mismatches 82; Indels 8; Gaps 2
QY	32	SWN-----ETQYDNLN-----QGYAETLEPRASAKKCDSEVYTGSETEGIPVAVYLSDFS	83	
DB	73	SWNDFLFSDDLRSODPLNFVDITDTYFNRLERKAKKNDPSRGVIVGACTISYPVALAVMDFS	132	
QY	84	FLGSLSTVAVSVIRMKAIHRTATELKLPLVSPASGARMQEDNRAFAVMVGSITAAVORHR	143	
DB	133	FMAGSGAVNGEKLTRILEKRIIDRLPIVYISASGARMQDSVPSLMQMAKTSALAKLH	192	
QY	144	EAHLPLVLYLNPTPTMGAMASWSSGHLTPAEPGAQIGFLGPRVVELTTGHALPDGVOQA	203	
DB	193	EAKLPVYSVLTNPTSGVATSPASLIGPVITAEPRALICFAGPRVVSQVIGDLPEGAQKS	252	
QY	204	ENLVKTVGIVGIVSPDLRAAVAKTL	229	
DB	253	EPFLHEGMIDKVERKQKTLTLESIL	278	

RESULT 15

ABU50977 standard; Protein; 279 AA.

AC ABU50977;

DT 08-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) protein #320.

KW Protein-protein interaction; ulcer; selected interacting domain; SID.

OS Helicobacter pylori.

PN WO200266501-A2.

PD 29-AUG-2002.

PF 28-DEC-2001; 2001WO-EP15428.

PR 02-JAN-2001; 2001US-259302P.

PA (HYBR-) HYBRIGENICS.
(INSP) INST PASTEUR.

PI Legrain P, Rain J, Coland F, De Reuse H, Labigne A;

DR WPI; 2002-674910/72.

DR N-PSDB; ABX65721.

PT New complexes of protein-protein interactions in Helicobacter pylori,
PT useful for identifying modulating compounds for treating or preventing
PT ulcers in mammals -

PS Claim 6; Page 164; 642pp; English.

CC The invention describes a complex of protein-protein interactions in
CC Helicobacter pylori selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful
CC for screening for agents which modulate the interaction of proteins.
CC Modulating compounds which binds to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This
CC is the amino acid sequence of a selected interacting domain (SID),
CC identified via protein-protein interactions.
CC Note: Where the patent number printed at the top of the pages in the
CC specification has obscured areas of protein sequence, the indexer
CC has replaced the residue with an X to represent an illegible residue.

SQ Sequence 279 AA;

Query Match 14.8%; Score 362; DB 23; Length 279;

Best Local Similarity 39.6%; Pred. No. 1.5e-24;

Matches 89; Conservative 40; Mismatches 80; Indels 16; Gaps 5;

QY 21 IDSVLDPPDSFISWNE--TPQYDNIN---QGYAETLERARSKACDESVITGEGTVEGIP 74

DB 56 IEFLLDVGSFEEFDGRLRN--DPLNXXVDKESYKQRIKKYEKRTNRSSVISGXAKINRMP 114

QY 75 VAVIISDFSGFLGSGITVASVIMKAIIRATLKLPLVSPASGARMQEDNRFAFMMVS 134

DB 115 LQ-IFWDSFWMGSLGSVEXEKIVRAINRAVAKREALLIVSASGARMOESTYSIMQMAK 173

QY 135 ITAAVORHREAHLPFLVLYRNPTMGAGAMASGSGHLTFAPBPAQIGFLGPRVVELTGH 194

DB 174 TSAALNRLEAKLPFLSLSLSDPTGGVSGSFAFLGDLIAEPGAMIGFAGPRVIVKOTIGA 233

QY 195 ALPDGVQAENLVKTGVIDGIVSPQLRAVA-----KTLKV 231

DB 234 DLPEGFQTAFLLEHGLIDMIYHRKDLKTLSDLIAMTHTKTSKI 278

Search completed: November 13, 2003, 06:40:21
Job time : 80 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:17:07 ; Search time 38 seconds

(without alignments)
1242.601 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448
Sequence: 1 VEKREPTVMWGMEHTSALT.....NNPERAGDRSFRFRRLAQ 491

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1047	42.8	495	2	E70783
2	431.5	17.6	301	2	AE3504
3	392.5	16.0	316	2	AE2101
4	384	15.8	289	2	E71875
5	384	15.7	289	2	F64638
6	375	15.3	304	1	AD0802
7	374	15.3	304	1	XMECBD
8	374	15.3	304	2	AE5873
9	374	15.3	304	2	H91028
10	372.5	15.2	284	2	A75422
11	370.5	15.1	282	2	F84045
12	369.5	15.1	276	2	F70340
13	368	15.0	288	2	C86722
14	368	15.0	296	1	B64113
15	363	14.8	262	2	G70001
16	361.5	14.8	298	2	AE2579
17	361.5	14.8	298	2	D97361
18	361.5	14.8	308	2	B71534
19	361.5	14.8	308	2	C66498
20	361.5	14.8	308	2	C72124
21	361	14.7	283	2	H81687
22	361	14.7	411	2	T07261
23	359.5	14.7	326	1	JQ1238
24	355.5	14.5	280	2	G81429
25	346.5	14.2	290	2	E81171
26	345.5	14.1	290	2	C81934
27	345.5	14.1	304	2	AH0337
28	344	14.1	317	2	AD3003
29	344	14.1	317	2	E98280

30	343.5	14.0	285	2	A89954	acetyl-CoA carboxy
31	336	13.7	288	2	A97338	acetyl-CoA carboxy
32	334	13.6	288	2	D95049	hypothetical prote
33	334	13.6	288	2	B97920	acetyl-CoA carboxy
34	334	13.6	308	2	B82253	acetyl-CoA carboxy
35	332.5	13.6	432	2	T06341	acetyl-CoA carboxy
36	332	13.6	307	2	D87688	hypothetical prote
37	331.5	13.5	489	2	S65564	acetyl CoA carboxy
38	331.5	13.5	507	2	T07012	acetyl-CoA carboxy
39	330.5	13.5	294	2	AG1633	acetyl-CoA carboxy
40	330.5	13.5	515	2	C72344	propionyl-CoA carb
41	329.5	13.5	290	2	C83257	acetyl-CoA carboxy
42	329	13.4	288	2	S73119	acetyl-CoA carboxy
43	327	13.4	516	2	A69527	methylnonyl-CoA
44	326.5	13.3	512	2	A05196	hypothetical prote
45	323.5	13.2	493	2	S78384	acetyl-CoA carboxy

ALIGNMENTS

RESULT 1									
E70783									
probable acd3 protein - Mycobacterium tuberculosis (strain H37Rv)									
C:Species: Mycobacterium tuberculosis									
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000									
C:Accession: E70783									
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.									
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.									
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.									
Nature 393, 537-544, 1998									
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.									
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome									
A:Reference number: AV0500; MUID:96295987; PMID:9634230									
A:Status: preliminary; nucleic acid sequence not shown; translation not shown									
A:Molecule type: DNA									
A:Residues: 1-495 <COI>									
A:Cross-references: GB:273101; GB:AL123456; NID:93261565; PIDN:CAA97379.1; PID:G1314044									
A:Experimental source: strain H37Rv									
C:Genetics:									
A:Gene: acd3									
C:Superfamily: propionyl-CoA carboxylase beta chain									
Query Match									
Best Local Similarity 42.8%; Score 1047; DB 2; Length 495;									
Matches 233; Conservative 73; Mismatches 153; Indels 28; Gaps 8;									
QY	23	SVLPDPSFISWNETPOYDNLNGVAFTELRASRAKODESVITGEGVBPVAVIISDF	82						
DB	12	AVLDKGSFVSWDEPLAVPVADSYARELAARATGADSVQTBGRVGRRAVAVACEF	71						
QY	83	SFLGSLGTVAIVRIMKAIHRATEIKLPLVSPASGARMQEDNRAFVMTVITAAVQRH	142						
DB	72	DFLGSGIGVAAARITAAVERATARPLPLASPSGGTRMGVAFQWTKIAAIOIOLH	131						
QY	143	REAHLPFLVYIRNPTMGAMASGSGHLTPAEFQAQIGPLGPVVEITTHALPDGVQO	202						
DB	132	NQARLPVLYVLRHPTTGGVFASWCSLGHITVAEFGALIGFGLPRVYELLYDPPPSGVQT	191						
QY	203	AENLVKGVINDIGIYSPLOBRAVAVKTLKVI-----QPVEATNRFSPPTPGVA--LPVMEAI	256						
DB	192	AENLRHGTIDGVVALDLRPMRLRALVLLDADPELPA-----PQTPAPVPDVPPTMSV	246						
QY	257	ARSRDPQRFGIGETMETLGAADVVLKSGARAGALSAPAVVALARIGRPVVLIGDRRP--	314						
DB	247	VASRRPDRPGVGRQLLRGATDRVLVLSGTDGEAATTL-LALARRGGQPTVLGGQRAVCG	305						
QY	315	---TLGPOELRPARGISLARLNLPIVSIITDSGAELSQAAEELGIASSIARTLSKIID	371						
DB	306	GGSTVGPALVREARGMALAEELCLPLVLYDAAGPALSAAEQGLAGQIAHCLAEIVT	365						
QY	372	APLPTVSIIGGVGGGLMLPADLVYAENNALSLPPEGAGAILPFDINHAETIER	431						

Db 366 LDTFVSTILGGSGGFPALMPADRVLANHGMPLPPEEGSALVFRDTAAAEIAAA 425
 QY 432 QGVQAHALLSQGLIDGIVAEETHFVEELG--TISNALSELDN-----NPERGRDSR 482
 Db 426 QGIRFADILKSGIVDTIYPEYFDPADEPIEFALRLSNATAAEVHALRKTIAPAPERLA--TR 483
 QY 483 FTRFEEL 489
 Db 484 LQRYRRI 490

RESULT 2

acetyl-CoA carboxylase [BC 6.4.1.2] [Imported] - *Bruceella melitensis* (strain 16M)
C:Species: *Bruceella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AF3504
R:DelVecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Majer, C.; Los, T.; Ivanova,
: Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hegius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Bruceella melitensis*
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <KUR>
A:Cross-references: GS:AE008917, PIDN:AAL53201.1; PTD:G17984075; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI2020
A:Map position: I
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain
C:Keywords: ligase

Query Match	17.6%	Score 431.5	DB 2	Length 301
Best Local Similarity	43.3%	Pred. No. 1e-15		
Matches 88	Conservative 43	Mismatches 69	Indels 3	Gaps 2

QY	46	YAEETTERASXXKCBESVITGETEYVIGIPAAVLISPSFSGSLGVAVSRIKALHPRAT	105
Db	93	YIDRLKDVASRRGMDDAIVNGJLSTIEGLPIVATVODPSFGSGSLGMGABEAILIQGEKAI	152
QY	106	EIKLPLTVSPASGAGAMQEDNRAFMVMSIITAACQRRREAHLPFLVYLINPTWGMASAS	165
Db	153	EIKRPLVFLPSSSGAGAMQEGISLMDLPRTTVAVENTMKKGLPIYIVLNPITGGTASV	212
QY	166	GSSGHLTPAEPGAQIGFLGPRVVELTTHGALPDGVQOAEMLYKTVIIGIVSPLQIPAAV	225
Db	213	AMLGDHIAEPGALIGFAGPRVLEQIIRERLPEFGQSSEYIMHGMDVMSVSLKATI	272
QY	226	AKTLKVI--QPVVATDRFSPTTP	246
Db	273	ARLLKIMTKOPAN--SDAPAFQKP	294

RESULT 3

acetyl-CoA carboxylase beta chain [imported] - Nostoc sp. (strain PCC 7120)
CisPeptides: Nostoc sp. PCC 7120
ANote: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C.Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C.Accession: AE2101
R.Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguchi, S.; Kuroda, M.; Tanaka, K.; Nakamura, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA.RS. 6, 205-213, 2001
Article: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A.Reference number: AB1807; MUID:21595285; PMID:11759840

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KUR>
A:Cross-references: GB:BA000019, PIDN:BAW74063.1, PID:gi17134456, GSPDB:GN00179

A; Experimental source: strain PCC 7120

C; Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match	16.0%;	Score 392.5;	DB 2;	Length 316;
Best Local Similarity	38.0%;	Pred. No. 2.8e-17;		
Matches 87; Conservative	42;	Mismatches 81;	Indels 19;	Gaps 3;

QY 32 SWNETPQYDNL-----NÖGYAETLERARSKAKCDESIVITGEGTVEGIPVAVILSD 81

Db 85 TWR--PMDENLRATDPLQFRDRKAYS DRLREMEDKLGLLDAVKTGLGQINSSPVALAVMD 14:

82 F5FLGSLGTVASVRIMKAIHRATELKLPLVSPASGGARMQEDNRAFMVMSITAAVQR 14.

Db 143 FRFMGSMGSVGEKITRLIEQATQRRYPVVICTSGGARMQEGMLSLMQAKISALER 20

Qy 142 HREAHLPLVYLNRPTMGAMASWGSSGHLTFAEPGAQIGFLGPRVELTTGHALPDGVQ 20

Db 203 HRDARLLYIPVLTNPPTGGVTASFAMLGDIILAEPKATIGFAGRRIEQLREKL.PDDFQ 26:

202 Q A E N L V K T G V I D G I V S P L Q L R A A V A K T L K V I Q E V E A T D R F S P T T P G V A L 250

Db 263 TAEDLLKHGFVDDIVPRTQLKNTLSQLIALHQPV-----PTTPMVL 304

RESULT 4

acetyl-coenzyme A carboxylase chain B - *Helicobacter pylori* (strain J99)

A;Variety: strain J99

C/Accession: E71875

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho

A;Accession: E71875

A;Molecule type: DNA

A;Cross-references: GB:AE001518;

Experimental Source: Bennett, 1995

C; Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.8%; Score 387.5; DB 2; Length 289;

Matches	91;	Conservative	43;	Mismatches	81;	Indels	15;	Gaps	4.
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

16 SÄTLIDSVLDPDSFISWNE--TPQYDNLN---QGYAETLERARSKAKCDESVITGEGT

Db 60 NATERIEFLCDVGSFEEFDKHLRPN-DPLNFVDKESYKQRIKKYEKRTNRPSVIGAEK 118

70 VEGIPVAVILSDFSFLGSLGTVASVRIMKAIHRATELKLPLLVSPASGGARMQEDNRAF 129

Db 119 INRMPLQIVVFDFSFMGSLGSEGEKIVRAINRAVAKKEALLIVSASGARMQESTYSL 178

QY 130 VMVSI^{TA}VQ^{RH}RE^{AH}L^{PL}V^{YL}R^{NP}TM^{GA}MA^{SW}SS^{GH}L^{TF}AE^{PG}AI^{GF}L^{GP}RV^E 189

Db 179 MQMAKTSALNRLSEAKLPFISLSDPTYGVSASFALGLLIAEPGAMIGFAGPRVIK 238

190 LTTGHALPDGVQDAENLVKTGVIDGIVSPQLRAVA-----KTLKV 231

Db 239 QITGADLPEGFQTAEFLLEHGLIDMIVHKDLKKTLSDLIAMTHKTSKI 288

RESULT 5

F64638
acetyl-CoA carboxylase beta subunit - *Helicobacter pylori* (strain 26695)

C:\species: helicobacter pylori

CISpecies: Escherichia coli
CISpecies: Escherichia coli
CIDate: 18-Jul-2001 #sequence 18-Jul-2001 #text_change 31-Dec-2001

Db 115 LHGLEVTLANVDFAFSGSGSMGVSVEEILARAADYAAESRTPILLVAASGGARMQESALSL 174

Qy 130 VMVSTAAVQRRHRAHLPVLYLRNPNGAMASWGSSGHLTFAPGAQIGLGRVVE 189
 Db 175 MQMAKTTVALMALSGRGGVYLSLTPTTGGVTAAPATADYIAIRGALLIGRGRVIG 234
 Qy 190 LTTGHALPQVQQAENLVKTGVIDGVSPQLRAAVALKLYIQPVEA 237
 Db 235 QTRONLPBGFQRAEFLLEHGMVDVNRREORRHLAGLGLVLTROEA 282

RESULT 11

P84045
 acetyl-CoA carboxylase transferase beta subunit accd [imported] - *Bacillus halodurans* (S
 C)Species: *Bacillus halodurans*
 C)Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C)Accession: F84045
 R)Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maeni, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A)Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A)Reference number: A83650; MUID:20512582; PMID:11058132
 A)Accession: F84045
 A)Status: preliminary
 A)Molecule type: DNA
 A)Residues: 1-282 <STO>
 A)Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BA06885.1; GSPDB:GN00
 A)Experimental source: strain C-125
 C)Genetics: accd
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.1%; Score 370.5; DB 2; Length 282;
 Best Local Similarity 36.5%; Pred. No. 5.7e-16;
 Matches 85; Conservative 42; Mismatches 85; Indels 21; Gaps 3;

Qy 11 GMEH-TSALTLDVLDPSFISMETPQYDLNOC-----YATIERARSKK 58
 Db 55 GFHHRMNAFDRLESLLDPGFV-----BLDKMTTEDEPLSPFYREKVEADRKTN 105
 Qy 59 CDESVITGEVTEGIPVAVILDSFSLGSLGVASVRIMKAIHRATELKLPLVSPASG 118
 Db 106 LNEALVTGEIGTNGFPVLYGVNDARFRKMSGVSVEKTRAIERAIENKQFPLPSAG 165
 Qy 119 GARMOEDNRFAVVMVSTAAVQRRHRAHLPVLYLRNPNGAMASWGSSGHLTFAPGA 178
 Db 166 GARMOEGVLSLQMAKTSALLERLDRVGLFISVMTHTPTTGGVSAFSLGQYNFAEPA 225
 Qy 179 QIGFGRVVELTTGHALPDGVQQAENLVKTGVIDGVSPQLRAAVALKLY 231
 Db 226 LIGFAGRRIEQTIREELPNDPQTAEFLKQGLDRVIRPSEMKDTLTILDI 278

RESULT 12

P70340
 acetyl-CoA carboxyltransferase beta subunit - *Aquifex aeolicus*
 C)Species: *Aquifex aeolicus*
 C)Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
 C)Accession: F70340
 R)Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy
 Nature 392, 353-358, 1998
 A)Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
 A)Reference number: A70300; MUID:98196666; PMID:9537320
 A)Accession: F70340
 A)Status: preliminary; nucleic acid sequence not shown; translation not shown
 A)Molecule type: DNA
 A)Residues: 1-216 <AQR>
 A)Cross-references: GB:AE000690; NID:g2983100; PIDN:AA06712.1; PID:g2983107; GB:AE00065
 A)Experimental source: strain VFS
 C)Genetics: accd
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain
 Query Match 15.1%; Score 369.5; DB 2; Length 276;
 Best Local Similarity 38.7%; Pred. No. 6.4e-16;

Matches 86; Conservative 42; Mismatches 89; Indels 5; Gaps 2;
 Qy 16 SALTLDVLDPS-FISMNTEPQYDNN-----OGYATLERARAKADESVINGEGV 70
 Db 50 SALERIKYTLDEENELLFEETLPADPLNFKDTSKYKRIKAQGETGISEAIVTEGHI 109
 Qy 71 EGIPIAVILDSFSLGSLGVASVRIMKAIHRATELKLPLVSPASGARMOEDNRFAV 130
 Db 110 KGRALIAVMDFNFIQSGMSGVSVERPFRACRAVAETPTLISFASGARMOEGISLM 169
 Qy 131 VMVSTAAVQRRHRAHLPVLYLRNPNGAMASWGSSGHLTFAPGAQIGLGRVVE 190
 Db 170 QMAKTTFGVGLKEANIRYISVLTNPNGVSAFAGLDLILAPKALIGRGRVIEQ 229
 Qy 191 TTHGHALPDGVQQAENLVKTGVIDGVSPQLRAAVALKLY 232
 Db 230 TTKQKLPEGFQYAEFLLEKQIDMVAHREKLDKLYALEVM 271

RESULT 13

C66722
 hypothetical protein accd [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)
 C)Species: *Lactococcus lactis* subsp. *lactis*
 C)Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C)Accession: C66722
 R)Bolotin, A.; Wincker, P.; Mauer, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
 Genome Res. 11, 731-753, 2001
 A)Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss
 A)Reference number: A86625; MUID:21235186; PMID:11337471
 A)Accession: C66722
 A)Status: preliminary
 A)Molecule type: DNA
 A)Residues: 1-288 <STO>
 A)Cross-references: GB:AE005176; PID:g12723698; PIDN:AA04877.1; GSPDB:GN00146
 A)Experimental source: strain IL1403
 C)Genetics: accd
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.0%; Score 368; DB 2; Length 288;
 Best Local Similarity 40.0%; Pred. No. 8.4e-16;
 Matches 86; Conservative 37; Mismatches 90; Indels 2; Gaps 2;

Qy 24 VLDPSFISMETPQYD-LN-OGYATLERARAKADESVINGEGVPAVILSD 81
 Db 72 VADKDSFVEMFTGIESKNPLDPGGIPKLAATKATGIDEAVITGATATKQKXTALATMD 131
 Qy 82 FSLGSLGVASVRIMKAIHRATELKLPLVSPASGARMOEDNRFAVVMVSTAAVOR 141
 Db 132 STFIAMSGIVGELTRLFLFVATTEKLPITVTRASGARMOEGISLMQMAKTSAAVYR 191
 Qy 142 HREAHLPVLYLRNPNGAMASWGSSGHLTFAPGAQIGLGRVVELTTGHALPDGVQ 201
 Db 192 HSNAGLFYITVLTPTTGGVTAAPATADYIAIRGALLIGRGRVIG 251
 Qy 202 QAENLVKTGVIDGVSPQLRAAVALKLYIQPVEA 236
 Db 252 KAELFNLHGFDAVAKRTLRQKLLALHLEHTEVE 286

RESULT 14

B64113
 acetyl-CoA carboxylase (EC 6.4.1.2) carboxyltransferase beta chain - *Haemophilus influen*
 C)Species: *Haemophilus influenzae*
 C)Date: 18-Aug-1995 #sequence_revision 23-Feb-1996 #text_change 18-Jun-1999
 C)Accession: B64113
 R)Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fritchman, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A)Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A)Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
 A)Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: B64113
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-236 <TIG>
A:Cross-references: GB:U32806; GB:L42023; NID:g1574717; PIDN:AAC2913.1; PID:g1574718; T
A:Note: named as homolog to a protein from *Escherichia coli*
C:Function:
A:Description: the enzyme complex catalyzes the synthesis of malonyl-CoA; the overall re
action bound to BCCP (see BKE9) with bicarbonate; carboxyltransferase catalyzes the trans
A:Pathway: fatty acid biosynthesis
A:Note: first committed step
A:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain
C:Keywords: fatty acid biosynthesis; ligase; zinc finger
F:29-31/Region: zinc finger CCCC motif

Query Match 15.0%; Score 368; DB 1; Length 296;
Best Local Similarity 36.8%; Pred. No. 8.7e-16; Indels 14; Gaps 3;
Matches 81; Conservative 45; Mismatches 80;

QY 25 LDDPSFISWNETPOY-DNLNQGVAETLERARSKACDESVITGEGVEGIPVAVILSDPS 83
DB 79 LEPRDLIKFDLKKYDRIINAQKETGEK-----DALITMTGTLNMPVIVAASNFA 130
QY 84 FLGSLGTVA SVRIMKAIRATELKLPLVSPASGARMQEDNRAFMVMSITRAVQRHR 143
DB 131 FMGSMGVSVAKFVKAERKAMENMCPFCFSASGARMOEALFSLQMAKTSAVLAQMR 190
QY 144 EAHLPFLVYLRNPTMGAMASWSSGHLTPAEPGAOIGFLGPRVELTGTALPDGVQA 203
DB 191 EKGVPFISVLDPLTGVSASFAMLDGINTAEFKALIGFAGPRVIEQTVREKLEPGFQRS 250
QY 204 ENLVKTVGIDIVSPLOLRAAVAKTLKVIQPEATDRFSP 243
DB 251 EFLEKGAIDMIVRSEMRQTASVLS-----KLITNPSP 285

RESULT 15

G70001
acetyl-CoA carboxylase homolog yctI - *Bacillus subtilis*

C:Species: *Bacillus subtilis*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: G70001
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, U.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G70001
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <KUN>
A:Cross-references: GB:Z99119; GB:AI009126; NID:g2635200; PIDN:CAB14881.1; PID:g2635386
A:Experimental source: strain 168
C:Genetics:
A:Gene: yctI
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 14.8%; Score 363; DB 2; Length 262;
Best Local Similarity 38.2%; Pred. No. 1.5e-15;
Matches 84; Conservative 41; Mismatches 89; Indels 6; Gaps 2;

QY 21 IDSVLDDPSFISWNETPOYDNL--NQGVAETLERARSKACDESVITGEGVEGIPVAVI 78

DB 38 IESIMDEGSFEEFQGMSENPLGPGTLEKLEKREKTSINBAVYCKGIGGHPAVYA 97
QY 79 LSDPSFLGSLGTVA SVRIMKAIRATELKLPLVSPASGARMQEDNRAFMVMSITRA 138
DB 98 VWDSSFRMGSMGVSVAKFVKAERKAMENMCPFCFSASGARMOEALFSLQMAKTS 157
QY 139 VORHREAHLPFLVYLRNPTMGAMASWSSGHLTPAEPGAOIGFLGPRVELTGTALPD 198
DB 158 IKLFSBEOGLIISVMTHTPTTGVSASFASLDGVNFAEPGALIGFAGRNRIIEQTI 217
QY 199 GVQAEENVKTVGIDIVSPLOLRAAVAKTLKVIQPEAT 238
DB 218 DQTAFLFKHQDLDAVH----RDMKKTLENLDMQGT 253

Search completed: November 13, 2003, 06:45:04
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 02:26:42 / Search time 24 Seconds

(without alignments)
962.088 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKRPPTVWGMEHTSALT.....NNPERAGRSRFRERLAQ 491

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1047	42.8	495	1	ACCD MYCTU Q10561 mycobacteri
2	389	15.9	305	1	ACCD SYN7 Q54776 synchococc
3	374	15.3	304	1	ACCD ECOLI P08193 escherichia
4	368	15.0	296	1	ACCD HAEIN P43778 haemophilus
5	361	14.7	411	1	ACCD CHIVU P56293 chlorella v
6	359.5	14.7	326	1	ACCD SYN3 Q57417 synchocyst
7	352.5	13.6	432	1	ACCD SOYBN P49158 glycine max
8	331.5	13.5	489	1	ACCD BRANA P48937 brassica na
9	330.5	13.5	501	1	ACCD LOTVA Q9bbs1 lotus japon
10	329	13.4	288	1	ACCD PORPU P51198 porphyra pu
11	326.5	13.3	512	1	ACCD TOBAC P12219 nicotiana t
12	324.5	13.3	488	1	ACCD ARATH P56765 arabidopsis
13	323.5	13.2	493	1	ACCD EPIV1 P30064 epifagus vi
14	322	13.2	314	1	ACCD ANGLY Q9m317 spinacia ol
15	321.5	13.1	522	1	ACCD SPIOL P52769 pinus thunb
16	318.5	13.0	321	1	ACCD PINTH P18823 pinus sativ
17	316.5	12.9	590	1	ACCD BEA P12217 marchantia
18	315	12.9	316	1	ACCD MARPO P13562 cucurbita ref
19	313.5	12.8	497	1	ACCD CUSRE Q9m613 oenothera h
20	311.5	12.7	438	1	ACCD OENHO Q31796 oenotheros
21	310	12.7	313	1	ACCD ANFTO P54541 bacillus su
22	302	12.3	506	1	ACCD BACSV O47039 picea abies
23	300	12.3	319	1	ACCD PICAB P48872 haemophilus
24	288	11.8	315	1	ACCD HAEIN Q00761 physcomitre
25	284	11.6	315	1	ACCD PHYP4 P53003 saccharopol
26	283.5	11.6	546	1	PCCB SACER O19903 cyanidium c
27	278.5	11.4	324	1	ACCA CYACA P40674 salmoneilla
28	269	11.0	318	1	ACCA SALTY P30867 escherichia
29	266	10.9	318	1	ACCA ECOLI P03166 homo sapien
30	264	10.8	539	1	PCCB HUMAN O67260 aquilex aeo
31	261.5	10.7	323	1	ACCA AQUAE P73384 sus scrofa
32	261	10.7	539	1	PCCB PIG O34847 bacillus su
33	260.5	10.6	325	1	ACCA_BACSV

34	259.5	10.6	326	1	ACCA SYN3 P74638 synchocyst
35	257	10.5	312	1	ACCA HELPU O92153 helicobacte
36	257	10.5	312	1	ACCA HELPU O25283 helicobacte
37	254	10.4	541	1	PCCB RAT P07633 rattus norv
38	252.5	10.3	327	1	ACCA SYN7 Q54766 synchococc
39	251	10.3	548	1	PCCS MYCTU P96885 mycobacteri
40	248.5	10.2	322	1	ACCA ANTSP P46316 antilhamio
41	246.5	10.1	549	1	PCCB MYCTE P53002 mycobacteri
42	238	9.7	324	1	ACCA PORPU P51371 porphyra pu
43	215.5	8.8	476	1	PCCB THOER O06101 rhodococcus
44	204.5	8.4	473	1	PCCB MYCTU Q10506 mycobacteri
45	190.5	7.8	563	1	MCCB_HUMAN Q9hcc0 homo sapien

ALIGNMENTS

RESULT 1
ACCD MYCTU STANDARD; PRT; 495 AA.
ID ACCD MYCTU
AC 010561:
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative acetyl-coenzyme A carboxylase subunit
DE beta (RC 6.4.1.2) (ACCase beta chain).
GN ACCD3 OR RV0904C OR MT0927 OR MTCY31.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Kiehn A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fletschman R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains." to the EMBL/GenBank/DBJ databases.
RL Submitted (Apr-2001)
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate + malonyl-CoA.
CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

Query	Match	Similarity	Score	DB	Length
Best Local	Similarity	47.8%	Pred. No. 1.1e-54		
Matches	233	Conservative	73	Missmatches	153
				Indels	28
				Gaps	8
DR	EMBL: Z73101; CA97379.1; -				
DR	AB006979; AAK45174.1; -				
DR	PIR: E70783; E70783.				
DR	TIGR; MT0927; -				
DR	Tuberclulst; RV0904c; -				
DR	InterPro; IPR000438; ACOACC_transfB.				
DR	InterPro; IPR000022; Carboxyl_trans.				
DR	Pfam; PF01039; Carboxyl_trans; 1.				
DR	PRINTS; PR01070; ACCCTRPRASEB.				
KM	Hypothetical protein; Fatty acid biosynthesis; Ligase;				
KM	Complete proteome.				
SQ	SEQUENCE 495 AA; 51772 MW; 15C96E921FF950C CR664;				
QY	23 SYLDPDSTISNMTQYDNINQVAFETLERASKXKCCESVYTGSGVIGIVAVILSDF				
Db	12 AVLDRGSFVSDSEPLAVPVADSVARELAAAPAGADESVITGSCRVRGRVAVACGF				
QY	83 SFLGSLGTVASVRIIMKAIHRATEKLPLVSPASGARMQEDNRAFVMVVSITAAQVR				
Db	72 DFLGSGIVAAAEIRITTAVERATAREPLASPSSGCTRMQGVAFLOMVKIAAIIQLH				
QY	143 REAHLPLVYLINPLTMGAMASWSSGHLTPAPQAIQFLGPRVVELTGHALPQNGQ				
Db	132 NOARLPVLYLHNPPTGVAFASWGSIGHLTVABPGALIGFLPVRVELLYGDPFPGVOT				
QY	203 AENLKGTGVYDIAVSEPLQRAAVAKTLKYI---QPEVATDRFSPTTPGVA--LPWMAI				
Db	192 AENLRHGIIDGVVALDRIRPMLDRALTYLIDAPELPA-----PQTPAPVPDPPTWDSV				
QY	257 AKSRDPQRPQIGELMETTGADVYKLSGAPAGALSPAVRYALAIRGERPVYLIGQDRF--				
Db	247 VASRRPDRPQVQLRLHGHGATDRLVLSGTQOGEAAITL-LALAFGGQPTVYVIGQGRVAVG				
QY	315 ---TGPOELRPARRGISLARELNPVSIIDPSGAELOAELGIASSIARTLSKIID				
Db	306 GGSIVGPAALRERRRMALAAELCLPLVLYIDAGALSAAPQGLAGIACHIAELVY				
QY	372 APLPTVSVIIGGVGSGALAMPADLVYAENAMLSALPEPGASAILFRDTNHAETIER				
Db	366 LDTPTVSIILGGSGGSPALAMLPADRVLAALHGMALPPEPGASAIIVFRDTAAAEALAA				
QY	432 QGVQAAHALISGLIDGIVAFETFEVEILG---TISNALSELDN-----NPERAGRDSR				
Db	426 QGIRSDALIKSGIVDTVPEYPPDADEPIEFALRLSNALAAVHALAKIIPAPERLA--TR				
QY	483 FTRFERL 489				
Db	484 LQRIYRI 490				
RESULT 2					
ID	ACCD_SYNYP7	STANDARD;	PRT;	305 AA.	
AC	OS4776;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Acetyl-coenzyme A carboxylase subunit beta				
DE	(EC 6.4.1.1.2) (ACMSE beta chain).				
GN	ACCD				
OS	Synechococcus sp. (strain PCC 7942) (Anacyctis nidulans R2).				
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.				
OX	NCBI TaxId=1140;				
RA	SEQUENCE FROM N.A.				
RA	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A				
CC	CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE				

CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANS-CARBOXYLASE
CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC + malonyl-CoA.
CC
CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHexamER OF BIOTIN
CC CARBOXYL CARRIER PROTEIN, BIOTIN COMPLEX (BY SIMILARITY).
CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U59237; AAB82034.1; -.
DR InterPro; IPR000438; AcOACC.transfB.
DR InterPro; IPR000022; Carboxyl_trans.1.
DR Pfam; PF01039; Carboxyl_trans.1.
DR PRINTS; PR01070; ACCCTFRPRASEB.
DR TIGRPFAMS; TIGR00515; acccd; 1.
KM Fatty acid biosynthesis; Ligase; Zinc-finger.
FT ZN_PING 33 55 C4-TYPE (POTENTIAL).
SQ SEQUENCE 305 AA; 33807 MW; 9AFCEFA5AE89F3E5A CAC64;

Query Match	15.9%;	Score 389;	DB 1;	Length 305;
Best Local Similarity	35.5%;	Pred. No. 3,1e-16;		
Matches	86;	Conservative	50;	Mismatches 90;
			Indels	16;
			Gaps	2

QY	21	IDSVDPSFTSMNETPOYDNLNCGYATLEERASKACDSVITTEGTGVSIPVAVLLS	80
DB	79	IDEAASFPDPPGFRD-----RKSYSDRLKETQANTGLSDAVRGVGLLEQPVALLGM	131
QY	81	DESFLGSGSLGVNASVRIMKAIHRAITELKPLTVSPASGARNQEDNRAFVMMVNSITAAVQ	140
DB	132	DPREFGSGMSGVGEKKTRLLEKGTQKSPVIIVYASGARNQEGMLSLMQAKISGLAE	191
QY	141	RHREAHLEFLVYLNRPTWGGAMASWSSGHLTPAEPGAQIGFLGRVVELTTGHALPDGV	200
DB	192	RHREGLLVLPILHTPTTGGVYASPMGLDIIIAEPKALIGFAGRRVIEQTLREKLPDGF	251
QY	201	QQAENVKTGYIDGVSPLOEBAVAVKLTXYQPEABDRSEPTTPGVALPWEALARS	260
DB	252	QTAELTQAHGFVDITVPRPTOLKKTLLAQLRLHQ-----PQSPMKPLLESSPAT	302
QY	261	DP 262	
DB	303	AP 304	

RESULT 3			
ACCD_ECOLI			
ID_ACCD_ECOLI	STANDARD;	PRT;	304 AA.
AC	P08193; P76937; P78251;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (EC 6.4.1.2) (ACCase beta chain).		
GN	ACCD OR DEDB OR USG OR B2316 OR Z5578 OR EC93200.		
OS	Escherichia coli, and		
OS	Escherichia coli O157:H7.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_TaxID=562, 83334;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=87308246; PubMed=3040739;		
RA	Bognar A.L., Osborne C., Shane B.;		
RT	Primary structure of the Escherichia coli folC gene and its		
	folypolylglutamate synthetase-dihydrofolate synthetase product and		

RA regulation of expression by an upstream gene." ;
 RL J. Biol. Chem. 262:12337-12343 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=87308226; PubMed=3040734;
 RA Nonet M.L., Marvel C.C., Tolan D.R.;
 RT "The hcr-purF region of the Escherichia coli K-12 chromosome.
 RL J. Biol. Chem. 262:12209-12217 (1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12." ;
 RL Science 277:1453-1474 (1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alta H., Baba T., Hayaishi K., Inada T., Iono K.,
 RA Itochi T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampel G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Mada C.,
 RA Yamagata S., Horinouchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features." ;
 RL DNA Res. 4:91-113 (1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. II, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Groetbeck E.J., Davis N.W., Lim A., Dimianta B.T., Potomousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner P.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7." ;
 RL Nature 409:529-533 (2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156233; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Murata T., Tanaka M., Tobe T.,
 RA Kunara S., Shibata T., Hattori M., Shiga H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12." ;
 RL DNA Res. 8:11-22 (2001).
 RN [7]
 RP SEQUENCE OF 1-31 FROM N.A.
 RX MEDLINE=93123150; PubMed=1678242;
 RA Li S.-J., Cronan J.E. Jr.;
 RT "Growth rate regulation of Escherichia coli acetyl coenzyme A
 RT carboxylase, which catalyzes the first committed step of lipid
 RT biosynthesis." ;
 RL J. Bacteriol. 175:332-340 (1993).
 RN [8]
 RP SIMILARITY TO ZPPA.
 RX MEDLINE=91360091; PubMed=1886618;
 RA Nagano Y., Matsuno R., Sasaki Y.;
 RT "An essential gene of Escherichia coli that has sequence similarity
 RT to a chloroplast gene of unknown function." ;
 RL Mol. Genet. 228:62-64 (1991).
 RN

[illegible]

```

ACCD_HAEMIN STANDARD; PRT; 296 AA.
ID ACCD_HAEMIN
AC P43778;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD OR H11260.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
CC CARBOXYLASE COMPLEX. FIRST, BIOTIN CARBOXYLASE CATALYZES THE
CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.
CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHOMER OF BIOTIN
CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS
CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32806; AAC22913.1; -.
CC PIR; B64113; B64113.
CC TIGR; H11260; -.
DR InterPro; IPR000438; ACOACC_transfB.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF01039; Carboxyl_trans. 1.
DR PRINTS; PRO1070; ACCCTFRASEB.
DR TIGRFAMS; TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Complete proteome.
FT ZN FING 29
FT C4-TYPE (POTENTIAL).
SQ SEQUENCE 296 AA; 32635 MW; 3FEE316557FD4553 CRC64;

Query Match 15.0%; Score 368; DB 1; Length 296;
Best Local Similarity 36.8%; Pred. No. 5, 1e-15;
Matches 81; Conservative 45; Mismatches 80; Indels 14; Gaps 3;

```

```

DB 191 EKVGFISVLTDPFLGVSASFAMLGDLNTAEBKALIGFAGPRVIEQTVAREKLPBEGFORS 250
QY 204 ENLYKTVINDIGIVSPDLRAAVAKTLKVIOPVATDRFSP 243
DB 251 EFLLEKGAIDMIVKRSERQTLASVLS-----KLTNQPS 285

RESULT 5
ACCD_CHLYU STANDARD; PRT; 411 AA.
ID ACCD_CHLYU
AC P56293;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.
OS Chlorella vulgaris.
OC Chlorophyta; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki U., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Sugitara M.;
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorella vulgaris: the existence of genes possibly
RT involved in chloroplast division."
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001684; BA57908.1; -.
CC PIR; T07261; T07261.
DR InterPro; IPR000438; ACOACC_transfB.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF01039; Carboxyl_trans. 1.
DR PRINTS; PRO1070; ACCCTFRASEB.
DR TIGRFAMS; TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 36
FT C4-TYPE (POTENTIAL).
SQ SEQUENCE 411 AA; 46547 MW; 81EB041499087E79 CRC64;

Query Match 14.7%; Score 361; DB 1; Length 411;
Best Local Similarity 31.7%; Pred. No. 1, 9e-14;
Matches 102; Conservative 60; Mismatches 114; Indels 46; Gaps 7;

```

DB 185 QWAKISALHHCQAKLYISLTPPTGVTASFAMLDLLFAEPKALIGFAGRVIE 244
 QY 190 LITTHALPDGVOQAENLVKTGVIDGIVSFLQRAAVAKTKYIQ--PVEATPRF-----S 242
 DB 245 QTLQEQLPDDPDTALYHGLDLIVPSFLKQALSEFLTYLKAPLKEQGRIPYGERG 304
 QY 243 PTTGVALPVMFAIRSDPQRPGI-GEIMETGADVLSGARAGALSPAVVALARIQ 301
 DB 305 PLTKRREGQLRFLKSSKTPREYLIVNDKELG----- 338
 QY 302 GRPVVLIGQDRRFTLGPQELRF 323
 DB 339 -----FLGQTF--TYPEKLEF 354

RESULT 6

ACCD_SYNY3 STANDARD; PRT; 326 AA.
 ID ACCD_SYNY3
 AC 057417;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DB Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
 DE (EC 6.4.1.2) (ACCase beta chain).
 GN ACCD OR ZPPA OR SLO336.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92118327; PubMed=1368738;
 RA Ogura Y., Yoshida T., Nakamura Y., Takemura M., Oda K., Ohnuma K.;
 RT "Gene encoding a putative zinc finger protein in Synechocystis
 PCC6803.";
 RL Agric. Biol. Chem. 55:2259-2264 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugita T., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 64k to 92k of the genome.";
 RL DNA Res. 2:153-166 (1995).
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
 CARBOXYLASE COMPLEX. FIRST, BIOTIN CARBOXYLASE CATALYZES THE
 TRANSCARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
 TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
 + malonyl-CoA.
 CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHOMER OF BIOTIN
 CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS
 OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; D10004; BAA0893.1; -;
 DR EMBL; S77740; AAC60398.1; -;
 DR EMBL; D63999; BAA10092.1; -;
 DR PIR; J01238; J01238.
 DR InterPro; IPR000438; ACOACC_transf.
 DR InterPro; IPR000022; Carboxyl_trans.
 DR Pfam; PF01039; Carboxyl_trans.1.
 DR PRINTS; PR01070; ACCCTRFRASEB.
 DR TIGRPFMS; TIGR00515; accd; 1.
 KW Fatty acid biosynthesis; ligase; zinc-finger; Complete proteome.

FT ZN FING 36 58 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 326 AA; 36396 MW; 99177806671A85B CRC64;

Query Match 14.7%; Score 359.5; DB 1; Length 326;
 Best local similarity 36.4%; Pred. No. 1.8e-14;
 Matches 80; Conservative 43; Mismatches 86; Indels 11; Gaps 2;

QY 44 QGVATLERARSKKACDESVITGEGTGVAVVLSFPFLGSLGVASRIMKAIHR 103
 DB 98 KSYKDIRIDTQANDLTDAVQTGHRGLDGLPALGVMPFRFGSGWSGVGEKLCRLLEY 157
 QY 104 ATELKPLPLVSPASGARGOEEDNRAFMVAIVSITAQVRRREAHLPFLVYLNPPTGGAMA 163
 DB 158 ATLERLPVITICASGARGMOEGMSLMQMATISGLQNHREKLYITVLTHTPTGGVTA 217
 QY 164 SWSSGHLTFAEPGAQIFLGPVVELTTHALPDGVOQAENLVKTGVIDGIVSPLQRA 223
 DB 218 SFAMLDLILAEFKATIGFAGRVIEQTLREKLPDQFTSEYLHGHVDAIVPAPQLKR 277
 QY 224 AVAKTKYIQVENVATDRSPPTGVALPVMFAIRSDPQ 263
 DB 278 TLAQLISLHQP-----FYP-----ILPPLNADSNQVNP 306

RESULT 7

ACCD_SOYEN STANDARD; PRT; 432 AA.
 ID ACCD_SOYEN
 AC P49158;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
 DE (EC 6.4.1.2) (ACCase beta chain).
 GN ACCD.
 OS Glycine max (Soybean).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Reank; TISSUE=leaf;
 RA Reverdatto S.V., Bellinson V., Nielsen N.C.;
 RT "The psbI, accD, psal, ORF 203, ORF 151, ORF 103, ORF 229 and psal
 RT gene cluster in the chloroplast genome of soybean.";
 RL (in) Plant Gene Register PGR95-051.
 CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
 INVOLVED IN CHLOROPLAST CARBON METABOLISM
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
 + malonyl-CoA.
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U26948; AAA80643.1; -;
 DR PIR; T06341; T06341.
 DR InterPro; IPR000438; ACOACC_transf.
 DR InterPro; IPR000022; Carboxyl_trans.
 DR Pfam; PF01039; Carboxyl_trans.1.
 DR PRINTS; PR01070; ACCCTRFRASEB.
 DR TIGRPFMS; TIGR00515; accd; 1.
 KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.
 FT ZN FING 169 191 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 432 AA; 48857 MW; 484256A7496ABDD CRC64;

Query Match 13.6%; Score 332.5; DB 1; Length 432;

Best Local Similarity 39.3%; Pred. No. 9 7e-13;
Matches 81; Conservative 32; Mismatches 82; Indels 11; Gaps 4;

QY 25 IDPDSFISNNETPOYDNLNGYAETLERARSKAKCDESVITGEGTGVGIPAVIILDSF 84
DB 222 IDPIEFHS-EEBP-----YKDRIDSYQRTGTLTAVQTGTGQLGIPVAILGMDPQF 272
QY 85 LGGSGTAVASVIMKAIHRATELKLPLVSPASGARMQEDNRAVVMVSTIAVQHRRE 144
DB 273 WGGSGASVAGEKITLVEYATNQLPLLVCSGARMQEGSLIDMAKISSALYDYQK 332
QY 145 AHLPLV-YLRNPITMGAMASWSSGHLTFAPGAQIGFIPRVYELTTHALPDGVOA 203
DB 333 NKKLVYSILTPTGGTAVTSGMGLDITIAFPNAVIAFAGKRVIEQTLNKAVPESQQA 392
QY 204 ENLVKTGVIDGIVSPQLQRAAVAKTL 229
DB 393 EYLFHKGJFDSIV-PRNLKGVLSL 417

RESULT 8

ACCD_BRANA STANDARD; PRT; 489 AA.

AC P48937;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.

OS Brassica napus (Rape).

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Brassica.

OX NCBI_TaxID=3708;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Jet neuf;
RX MEDLINE=96207569; PubMed=8670092;
RA Elborough K.M., Wanz R., Deke R.K., Markham J.E., White A.J.,
RA Rawsthorne S., Slabas A.R.;

RT "Biotin carboxyl carrier protein and carboxyltransferase subunits of
RT the multi-subunit form of acetyl-CoA carboxylase from Brassica napus:
RT cloning and analysis of expression during oilseed rape

RT embryogenesis.";
RL Biochem. J. 315:103-112(1996).
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.

CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL: Z50868; CA90747.1; -
CC PIR: S66564; S66564.
DR InterPro: IPR000438; ACOACC.transfb.
DR InterPro: IPR000022; Carboxyl.trans.
DR Pfam: PF01039; Carboxyl.trans.1.
DR PRINTS: PRO1070; ACCCTFRASEB.
DR TIGRFAMs: TIGR00515; accd.1.
KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.
FT ZN FING 229 248 C4-TYPE (POTENTIAL).
SQ SEQUENCE 489 AA; 55476 MW; 540C79FBA239432 CRC64;

Query Match 13.5%; Score 331.5; DB 1; Length 489;

Best Local Similarity 34.4%; Pred. No. 1.3e-12;
Matches 76; Conservative 50; Mismatches 82; Indels 13; Gaps 3;

QY 21 IDSVLPDPDSFISNN-----ETPOYDNLNGYAETLERARSKAKCDESVITGEGTVE 71
DB 260 IELSIDG-----SNWGMEDWVSADPIKFHREERYKRIASQKGTGLTAIGTQQLN 316
QY 72 GIPVAVILDSFISGSLGVASVIMKAIHRATELKLPLVSPASGARMQEDNRAVFM 131
DB 317 GIPVAVLMFOFQWGGSGVGDKITRLILEYATNQLPLLVCSGARMQEGSLIMQ 376
QY 132 MVSTIAVQHRRE-HLPVLVLENPITMGAMASWSSGHLTFAPGAQIGFIPRVYEL 190
DB 377 MAKISVLCIDIOSSKULPYISILTPTGGTAVTSLGMLGITIAEPYAVIAFAGKRVIEQ 436
QY 191 TTHALPDGVOAENLVKTGVIDGIVSPQLQRAAVAKTLV 231
DB 437 TLKRAVPEGQAESILRKGLDAIVPRNPLKGVVSLPQL 477

RESULT 9

ACCD_LOTUA STANDARD; PRT; 501 AA.

AC Q9BBS1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.

OS Lotus japonicus.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots I; Fabales; Fabaceae; Papilionoideae; Lotaseae; Lotus.

OX NCBI_TaxID=34305;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Accession MG-20;
RX MEDLINE=21082929; PubMed=11214967;
RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;

RT "Complete structure of the chloroplast genome of a legume, Lotus
RT japonicus.";
RL DNA Res. 7:323-330(2000).

CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.

CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL: AP002983; BAB3205.1; -
CC InterPro: IPR000438; ACOACC.transfb.
DR InterPro: IPR000022; Carboxyl.trans.
DR Pfam: PF01039; Carboxyl.trans.1.
DR PRINTS: PRO1070; ACCCTFRASEB.
DR TIGRFAMs: TIGR00515; accd.1.
KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.
FT ZN FING 235 257 C4-TYPE (POTENTIAL).
SQ SEQUENCE 501 AA; 57038 MW; C08BBB343D7DE54D CRC64;

Query Match 13.5%; Score 330.5; DB 1; Length 501;
Best Local Similarity 36.3%; Pred. No. 1.5e-12;
Matches 77; Conservative 42; Mismatches 78; Indels 15; Gaps 4;
QY 21 IDSVLPDPDSFISNN-----ETPOYDNLNGYAETLERARSKAKCDESVITGEGTVE 71

```

Db      269 IELSLIDPG---TWNPEDMDMISVDPIEFHSEEPYKDRIDSYOKTTGTLTEAVOTGCHLN 325
Qy      72 GIPVAVILSDPSFLGSLGTVAASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFMN 131
Db      336 GIPVAVIAGIMDFEFMGSGMSVGEKILTRLVEYATINOLPLIVCASGARMQEGSLSLMQ 385
Qy      132 MVSITAAVQRHR-EAHLPLVYLRLNPTMGAMASMGSGHLFPAPGAOIGFLGPRVETL 190
Db      386 MAKISSALYDYOINKKLFFVYSILSTPTTGSGVTRASFQMLGDIITAEFNATYARAKRIVIEQ 445
Qy      191 TTGHALPDGVQOAEMLVKTGVIDGIV--SPLQ 220
Db      446 TLNKAIVEGSGQAAYELFHKKLFDSDIYRNPPLK 477

RESULT 10
ACCD_PORPU STANDARD; PRT; 288 AA.
ID ACCD_PORPU STANDARD; PRT; 288 AA.
AC P51198;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
   (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxId=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munholand J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
   genome."
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
   INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
   + malonyl-CoA.
CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U38804; AAC08084.1; -
DR PIR; S73119; S73119.
DR InterPro; IPR000438; ACOACC.transFB.
DR InterPro; IPR000022; Carboxyl.trans.
DR Pfam; PF01039; Carboxyl.trans.1.
DR PRINTS; PRO1070; ACCCTRFRASEB.
DR TIGRfam; TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 34
FT SEQUENCE 288 AA; 32168 MW; C2FD4C8A04D8E261 CRC64;

Query Match 13.4%; Score 329; DB 1; Length 288;
Best Local Similarity 38.6%; Pred. No. 9.7e-13;
Matches 71; Conservative 39; Mismatches 74; Indels 0; Gaps 0;

```

```

Qy      166 GSSGHLTPAPGAOIGFLGPRVETLTGHALPDGVQOAEMLVKTGVIDGIVSPLQRAV 225
Db      218 AMLGDIITAEFPALIAFAGRRVIEQITIKEDLPDNQSSYLEFHEGFLDILVTRTQRSKL 277
Qy      226 AKTL 229
Db      278 IQIL 281

RESULT 11
ACCD_TOBAC STANDARD; PRT; 512 AA.
ID ACCD_TOBAC STANDARD; PRT; 512 AA.
AC P12219;
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
   (EC 6.4.1.2) (ACCase beta chain).
GN ACCD OR YCP11 OR ZPPA.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxId=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bright yellow 4;
RA Shinozaki K., Ohme M., Tanaka M., Wakauchi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torazawa K., Meng B.-Y., Sugita M.,
RA Dene H., Kamogashira T., Yamada K., Kusuda J., Takiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
   its gene organization and expression."
RL EMBO J. 5:2043-2049(1986).
CC -!- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
   INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
   + malonyl-CoA.
CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z00044; CAAT7362.1; -
DR PIR; A05196; A05196.
DR InterPro; IPR000438; ACOACC.transFB.
DR InterPro; IPR000022; Carboxyl.trans.
DR Pfam; PF01039; Carboxyl.trans.1.
DR PRINTS; PRO1070; ACCCTRFRASEB.
DR TIGRfam; TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 247
FT SEQUENCE 512 AA; 58472 MW; DFEC3132E6228868 CRC64;

Query Match 13.3%; Score 326.5; DB 1; Length 512;
Best Local Similarity 34.4%; Pred. No. 2.7e-12;
Matches 75; Conservative 49; Mismatches 87; Indels 7; Gaps 2;

```

QY 135 ITAAVQRHR-EAHLPLVLYRNPTMGAMASWSSGHLTPAEPGAQIGFLGRVVELTTG 193
 DB 401 ISSALYDQVQNKKKLKYVVSILTSPTTGCVTASFGMLGDIITIAEPNAYIAFAGKRVIEQTLN 460
 QY 194 HALPDGVOAENLVKTVGIDIVSPLQTRAFAVAKTLKY 231
 DB 461 KTVPEGSOAEYLFQKGLFDLIVPRNLKLSVLSLFLK 498

RESULT 12

ACCD ARATH STANDARD; PRT; 488 AA.
 ID ACCD ARATH STANDARD; PRT; 488 AA.
 AC P56765;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
 DE (EC 6.4.1.2) (ACCase beta chain).
 GN ACCD OR ATCG00500.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Wen T.N., Wurtele E.S., Nikolaus B.J.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
 RX MEDLINE=20039611; PubMed=10574454;
 RT "Complete structure of the chloroplast genome of Arabidopsis
 thaliana.";
 RL DNA Res. 6:283-290(1999).

CC -1- INVOLVED IN PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
 CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
 CC + malonyl-CoA.
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF056971; AAF35256.1; -;
 CC DR EMBL; AP000423; BAA84394.1; -;
 CC DR InterPro; IPR000438; ACOACC transfb.
 CC DR InterPro; IPR000022; Carboxyl_trans.
 CC DR Pfam; PF01039; Carboxyl_trans; 1.
 CC DR PRINTS; PR01070; ACCCTRFRASEB.
 CC DR TIGRFBMS; TIGR00515; accd; 1.
 CC DR Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
 CC ZN FING 231 250 C4-TYPE (POTENTIAL).
 CC FT SEQUENCE 488 AA; 55609 MW; 389B7C0B213A71CE CRC64;
 CC SQ

Query Match 13.3%; Score 324.5; DB 1; Length 488;
 Best Local Similarity 33.5%; Pred. No. 3.3e-12;
 Matches 74; Conservative 52; Mismatches 82; Indels 13; Gaps 3;

QY 21 IDSVLDPDSFTISWN-----ETPQYDNLNGVAFLETRARSAKODESVITGEGTIVE 71
 DB 262 IELSLDPCG--TWNMEDDMVSADPIKHSKEPKNRIDSQKTTGTLDVAGTGTGQLN 318
 QY 72 GIPAVILSDPSFLGSLGTVASVRIMKAIRHATELKLPLVSPASGARGMDEDRFAVM 131

DB 319 GIPALGVMPDRFWMGSMGVGEKITRLIEVATNQCLPLILVSSCGARMQEGSLIMQ 378
 QY 132 MVSTIAVQRHRE-HLPELVLYRNPTMGAMASWSSGHLTPAEPGAQIGFLGRVVEL 190
 DB 379 MAKISSVLCDYQSSKKLKYVVSILTSPTTGCVTASFGMLGDIITIAEPNAYIAFAGKRVIEQ 438
 QY 191 TTHALPDGVOAENLVKTVGIDIVSPLQTRAFAVAKTLKY 231
 DB 439 TLKKAIVEGSOAEYLFQKGLFDLIVPRNLKLSVLSLFLK 479

RESULT 13

ACCD EPIVI STANDARD; PRT; 493 AA.
 ID ACCD EPIVI STANDARD; PRT; 493 AA.
 AC P30054;
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
 DE (EC 6.4.1.2) (ACCase beta chain).
 GN ACCD OR YCF1 OR ZPPA OR DEDB.
 OS Epifagus virginiana (Beechdrops).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Lamiales; Orobanchaceae; Orobanchaeae; Epifagus.
 OC NCBI_TaxID=4177;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93066301; PubMed=1332054;
 RA Wolfe K.H., Morden C.W., Palmer J.D.;
 RT "Function and evolution of a minimal plastid genome from a
 RT nonphotosynthetic parasitic plant";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652(1992).
 CC -1- INVOLVED IN PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
 CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
 CC + malonyl-CoA.
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M81884; AAA65854.1; -;
 CC DR PIR; S78384; S78384.
 CC DR InterPro; IPR000438; ACOACC transfb.
 CC DR InterPro; IPR000022; Carboxyl_trans.
 CC DR Pfam; PF01039; Carboxyl_trans; 1.
 CC DR PRINTS; PR01070; ACCCTRFRASEB.
 CC DR TIGRFBMS; TIGR00515; accd; 1.
 CC DR Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
 CC ZN FING 235 257 C4-TYPE (POTENTIAL).
 CC FT SEQUENCE 493 AA; 56560 MW; ZCB0529471558036 CRC64;
 CC SQ

Query Match 13.2%; Score 323.5; DB 1; Length 493;
 Best Local Similarity 33.5%; Pred. No. 3.8e-12;
 Matches 73; Conservative 51; Mismatches 87; Indels 7; Gaps 2;

QY 21 IDSVLDPDSFTISWNETP-----QYDNLNGVAFLETRARSAKODESVITGEGTIVE 74
 DB 269 IEVLVDDDTWYPMEDWSSLDPIEFHSEEPYDRIVSYOKRTGLTEAVQTGTGQLNGIP 328
 QY 75 VAVILSDPSFLGSLGTVASVRIMKAIRHATELKLPLVSPASGARGMDEDRFAVMVS 134
 DB 329 IALGVMPFQFMGSMGVGEKITRLIEVATNKILPLIVCASGARMQEGSLIMQMAK 388
 QY 135 ITAAVQRHRE-EAHLPLVLYRNPTMGAMASWSSGHLTPAEPGAQIGFLGRVVELTTG 193

Db 389 ISSALFDYOSNNKLLYVSLTPTTGGVYASFGMLDIIIAEBSYIAFAKRVIEQTLH 448
194 HALPDGVOAENLVKTGVIDGIVSPQLRAAVAKTLKV 231
449 KIVPESQAAEYLFQKGLFDLIIPNNLKSJVLGELFKL 486

RESULT 14

ACCD ANGLY

ID ACCD ANGLY STANDARD; PRT; 314 AA.

AC P28252;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
(EC 6.4.1.2) (ACCase beta chain).
GN ACCD OR YCF11 OR ZFPA OR DEDB.
OS Angiopteris lygodifolia (Turnip fern).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Filicophyta; Marattiopsida; Marattiales; Marattiaceae; Angiopteris.
OX NCBI_Taxid=3267;

RM SEQUENCE FROM N.A.
RP STRAIN=Rosenstock;
RX MEDLINE=92119238; PubMed=1731980;
RA Yoshinaga K., Kubota Y., Ishii T., Wada K.;
RT "Nucleotide sequence of atps, tbcL, trnR, deeb and psal chloroplast
genes from a fern Angiopteris lygodifolia: a possible emergence of
Spermatophyta lineage before the separation of Bryophyta and
Pteridophyta.";
RL Plant Mol. Biol. 18:79-82(1992).
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
+ malonyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X58429; CAA41333.1; -
DR PIR; S19230; BWFMT.
DR InterPro; IPR000438; ACOCAC transFB.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF01039; Carboxyl_trans; 1.
DR PRINTS; PR01070; ACCOTRFRASEB.
DR TIGRFAMs; TIGR00515; accd; 1.
DR Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 51 73 C4-TYPE (POTENTIAL).
SQ SEQUENCE 314 AA; 35098 MW; 0B09A7DF118D9D28 CRC64;

Query Match 13.2%; Score 322; DB 1; Length 314;
Best Local Similarity 34.1%; Pred. No. 2.8e-12;
Matches 74; Conservative 46; Mismatches 91; Indels 6; Gaps 2;

Qy 21 IDSVLDPSPFSIMNTPQYDNL-----NCGVAETLEBARSKAKCDESVITGGTVEGIPV 75
Db 85 IELLDRIWIPMDMDVAQDVLAQKSDSDSYNNRISLSQKRGTLDAVQTGIGNANGPVP 144
Qy 76 AVILSDPFLGSLGTVA SVIRIMKAIRHATELKLPLVSPASGARMOEDNRAFYMMVSI 135
Db 145 ALGVMDPFGSMGSMGVGKTRILIEVATQESLPLIIVCASGARMQEGITLSSQMOMKI 204
Qy 136 TAAVQRHR-EAHLPLVLVLRNPTMGAMASWGSSGHLTFAEQAQIGFLGPVVELTGGH 194
Db 205 SSVSQIHQVOKKLLYAVLTVPTTGGVYASFGMLDIIIAEBSYIAFAKRVIEQTLH 264

Qy 195 ALPDGVOAENLVKTGVIDGIVSPQLRAAVAKTLKV 231
Db 265 KIPDFQVAESLFDHGLDSTVPRNLKGVLSERTEL 301

RESULT 15

ACCD SPIOL

ID ACCD SPIOL STANDARD; PRT; 522 AA.

AC 09M3L7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
(EC 6.4.1.2) (ACCase beta chain).
GN ACCD.
OS Spinacia oleracea (Spinach).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_Taxid=33562;

RM SEQUENCE FROM N.A.
RP STRAIN=cv. Geant d'hiver, and cv. Monatol;
RX MEDLINE=21187424; PubMed=11292076;
RA Schnitz-Linneweber C., Maier R.M., Alcaraz J.-P., Cottet A.,
RA Hermann R.G., Maché R.;
RT "The plastid chromosome of spinach (Spinacia oleracea): complete
nucleotide sequence and gene organization.";
RL Plant Mol. Biol. 45:307-315(2001).
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
+ malonyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AJ400848; CAB88738.1; -
DR InterPro; IPR000438; ACOCAC transFB.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF01039; Carboxyl_trans; 1.
DR PRINTS; PR01070; ACCOTRFRASEB.
DR TIGRFAMs; TIGR00515; accd; 1.
DR Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 258 280 C4-TYPE (POTENTIAL).
SQ SEQUENCE 522 AA; 58907 MW; 84A337F76A3BF87 CRC64;

Query Match 13.1%; Score 321.5; DB 1; Length 522;
Best Local Similarity 34.1%; Pred. No. 5.3e-12;
Matches 76; Conservative 45; Mismatches 85; Indels 17; Gaps 4;

Qy 21 IDSVLDPSPFSIMNTPQYDNL-----NCGVAETLEBARSKAKCDESVITGGTVEGIPV 69
Db 22 IELLDIPG---TWN--PMDDDMVSNPDPIGFHSEEBAYCDRDSYQIKRGLEAVQTGIGQ 346
Qy 70 VEGIPVAVILSDPFLGSLGTVA SVIRIMKAIRHATELKLPLVSPASGARMOEDNRAFY 129
Db 347 LINGIPVAIGVMDPFGSMGSMGVGKTRILIEVATQESLPLIIVCASGARMQEGISTL 406
Qy 130 VMVMSITAAVQRHREAHLPVLV-YLRNPTMGAMASWGSSGHLTFAEQAQIGFLGPV 188
Db 407 MQMAKISSVLYDYOSNNKLLYVSLTPTTGGVYASFGMLDIIIAEBSYIAFAKRVIE 466
Qy 189 ELTTGHALPDGVOAENLVKTGVIDGIVSPQLRAAVAKTLKV 231

Db 467 EQTLNKTVPEGSGQAALFLFHKGFLPDPVIVRNILKGVLSLPEL 509

Search completed: November 13, 2003, 06:41:00
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 05:43:33 ; Search time 89 Seconds
(without alignments)
1423.639 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKREPTVMWGMHTSALT.....NNPERAGRSRFRERLAQ 491

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2445	99.9	491	16	Q93KF1
2	1527.5	62.4	483	16	Q93KF1
3	599.5	24.5	578	2	Q93KF1
4	480	19.6	458	16	Q93KF1
5	431.5	17.6	301	16	Q93KF1
6	431.5	17.6	301	16	Q93KF1
7	427	17.4	308	16	Q93KF1
8	409	16.7	1338	5	Q93KF1
9	404	16.5	272	2	Q93KF1
10	392.5	16.0	316	16	Q93KF1
11	389.5	15.9	304	16	Q93KF1
12	387.5	15.8	289	16	Q93KF1
13	386	15.8	309	16	Q93KF1
14	384	15.7	289	16	Q93KF1
15	379	15.5	201	2	Q93KF1
16	375	15.3	304	16	Q93KF1

17	374	15.3	332	16	Q93KF1	Q93KF1
18	372.5	15.2	284	16	Q93KF1	Q93KF1
19	370.5	15.1	282	16	Q93KF1	Q93KF1
20	369.5	15.1	276	16	Q93KF1	Q93KF1
21	368	15.0	288	16	Q93KF1	Q93KF1
22	367	15.0	304	16	Q93KF1	Q93KF1
23	365	14.9	291	16	Q93KF1	Q93KF1
24	365	14.9	291	16	Q93KF1	Q93KF1
25	363	14.8	262	16	Q93KF1	Q93KF1
26	361.5	14.8	298	16	Q93KF1	Q93KF1
27	361.5	14.8	308	16	Q93KF1	Q93KF1
28	361.5	14.8	308	16	Q93KF1	Q93KF1
29	361	14.7	283	16	Q93KF1	Q93KF1
30	360.5	14.7	288	16	Q93KF1	Q93KF1
31	359	14.7	290	16	Q93KF1	Q93KF1
32	355.5	14.5	280	16	Q93KF1	Q93KF1
33	354.5	14.5	285	16	Q93KF1	Q93KF1
34	346.5	14.2	290	16	Q93KF1	Q93KF1
35	345.5	14.1	290	16	Q93KF1	Q93KF1
36	345.5	14.1	304	16	Q93KF1	Q93KF1
37	345.5	14.1	356	16	Q93KF1	Q93KF1
38	344	14.1	317	16	Q93KF1	Q93KF1
39	343.5	14.0	285	16	Q93KF1	Q93KF1
40	343.5	14.0	285	16	Q93KF1	Q93KF1
41	343.5	14.0	304	16	Q93KF1	Q93KF1
42	343	14.0	288	16	Q93KF1	Q93KF1
43	343	14.0	288	16	Q93KF1	Q93KF1
44	339	13.8	297	8	Q93KF1	Q93KF1
45	338.5	13.8	295	16	Q93KF1	Q93KF1

ALIGNMENTS

RESULT 1	ID	Q93KF1	PRELIMINARY	PRT	491 AA.
AC	Q93KF1	01-DEC-2001 (TREMBL)	19, Last sequence update)		
DT	01-DEC-2001 (TREMBL)	19, Last sequence update)			
DT	01-MAR-2003 (TREMBL)	23, Last annotation update)			
DE	Putative carboxyltransferase subunit of acetyl-CoA carboxylase				
DE	(Acetyl-CoA carboxylase beta subunit) (EC 6.4.1.2).				
GN	ACCPA OR CGL0831.				
OS	Corynebacterium glutamicum (Brevibacterium flavum).				
OC	Bacteria; Actinobacteria; Actinobacteriales;				
OC	Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.				
OX	NCBI TaxID=1718;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 13032;				
RA	Kim K., Peters-Wendisch P.G., Sahn H., Birkmann B.J.;				
RL	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 13032 / DSM 20300 / NCIB 10025;				
RA	Nakagawa S.;				
RT	Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.				
RL	EMBL, Y17592; CAC42827.1; -				
DR	EMBL, Y17592; CAC42827.1; -				
DR	EMBL, AP005276; BAB98224.1; -				
DR	InterPro: IPR000438; ACOACC; trans.				
DR	InterPro: IPR000022; Carboxyl; trans.				
DR	InterPro: IPR001753; EnCoA_hyd; trans.				
DR	Pfam: PF01039; Carboxyl; trans.				
DR	Pfam: PF01039; ECH; 1.				
DR	PRINTS: PRO1070; ACCCTPRASEB.				
KW	ligase; Transferase; Complete proteome.				
SO	SEQUENCE 491 AA; 52238 MW; 53D1B875A92D82D CRC64;				

Query Match 99.9%; Score 2445; DB 16; Length 491;
Best Local Similarity 99.8%; Pred. No. 3.5e-138;
Matches 490; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 VEKRPETWVGMEHTSALTLLIDSVLPDPSFISMNTPQYDNLNOGYAETLERARAKACD 60
D 1 MEKRPETWVGMEHTSALTLLIDSVLPDPSFISMNTPQYDNLNOGYAETLERARAKACD 60
QY 61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGA 120
D 61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGA 120
QY 121 RMOENRNFVMMVSTTAAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQI 180
D 121 RMOENRNFVMMVSTTAAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQI 180
QY 181 GFELGRVVELTTHGALPDGVOAENLVKTGVIDGIVSPLOLRAAVALKVIQPEATDR 240
D 181 GFELGRVVELTTHGALPDGVOAENLVKTGVIDGIVSPLOLRAAVALKVIQPEATDR 240
QY 241 FSPPTPGVALPVMEAIARSRDPORPGIGIMETTLGADVVKLSGARAGALSPAVVALARI 300
D 241 FSPPTPGVALPVMEAIARSRDPORPGIGIMETTLGADVVKLSGARAGALSPAVVALARI 300
QY 301 GGRPVVILIGODRRFTLGPOELRPARRGISLARELMPLVSIIDTSGAEISQAABELGIAS 360
D 301 GGRPVVILIGODRRFTLGPOELRPARRGISLARELMPLVSIIDTSGAEISQAABELGIAS 360
QY 361 SIARTLSKLIDAPLPTVSITIGGVGGGALAMLPAADVAAENAMLSALPPEGASAILFR 420
D 361 SIARTLSKLIDAPLPTVSITIGGVGGGALAMLPAADVAAENAMLSALPPEGASAILFR 420
QY 421 DTNHAABIEROGVOAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRD 480
D 421 DTNHAABIEROGVOAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRD 480
QY 481 SRFTEFERLAQ 491
D 481 SRFTEFERLAQ 491
QY 481 SRFTEFERLAQ 491
D 481 SRFTEFERLAQ 491

RESULT 2
Q8FR58 PRELIMINARY; PRT; 483 AA.
ID Q8FR58
AC Q8FR58;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative carboxyltransferase subunit of acetyl-CoA carboxylase.
GN ACCDA OR CE0907.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005217; BAC17717.1; -
KW Transferase; Complete proteome.
SQ SEQUENCE 483 AA; 51545 MW; 653BA89519BBDC7 CRC64;
```

Query Match 62.4%; Score 1527.5; DB 16; Length 483;
Best Local Similarity 62.9%; Pred. No. 2.2e-83;
Matches 302; Conservative 62; Mismatches 115; Indels 1; Gaps 1;

```
QY 12 MEHTSALTLLIDSVLPDPSFISMNTPQYDNLNOGYAETLERARAKACDESITGEGTVE 71
D 12 MEHTSALTLLIDSVLPDPSFISMNTPQYDNLNOGYAETLERARAKACDESITGEGTVE 71
QY 1 MTRTSARDLIDVLDKDSFCFMDTPPEYGEIDDDYREALARAKRSQVDEAVITGEGTVE 60
D 1 MTRTSARDLIDVLDKDSFCFMDTPPEYGEIDDDYREALARAKRSQVDEAVITGEGTVE 60
QY 72 GIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMM 131
D 72 GIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMM 131
```

```
D 61 GSRVAFIVSEFAFLGSGIGATSRRLLIOAIRATRERLPLLISPGSGTRMOEGTAPFAM 120
QY 132 MVSTTAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELT 191
D 132 MVSTTAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELT 191
QY 121 MISTITAVRRKDHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELT 180
D 121 MISTITAVRRKDHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELT 180
QY 192 TGHALPDGVOAENLVKTGVIDGIVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 251
D 192 TGHALPDGVOAENLVKTGVIDGIVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 251
QY 252 VMEAIARSRDPORPGIGIMETTLGADVVKLSGARAGALSPAVVALARIGRPVVLIGOD 311
D 252 VMEAIARSRDPORPGIGIMETTLGADVVKLSGARAGALSPAVVALARIGRPVVLIGOD 311
QY 312 R-RTTLGPOELRPARRGISLARELMPLVSIIDTSGAEISQAABELGIASIAITLSKLI 370
D 312 R-RTTLGPOELRPARRGISLARELMPLVSIIDTSGAEISQAABELGIASIAITLSKLI 370
QY 371 DAPLPTVSITIGGVGGGALAMLPAADVAAENAMLSALPPEGASAILFRDTNHAABIE 430
D 371 DAPLPTVSITIGGVGGGALAMLPAADVAAENAMLSALPPEGASAILFRDTNHAABIE 430
QY 431 RQGVQAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRSRFTFRERLA 490
D 431 RQGVQAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRSRFTFRERLA 490
QY 490 RQGVQAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRSRFTFRERLA 490
D 490 RQGVQAHALLSGGLIDGIVAEETHEFVEEILGTISNALSELDDNPERAGRSRFTFRERLA 490
```

```
RESULT 3
Q9F6D6 PRELIMINARY; PRT; 578 AA.
ID Q9F6D6
AC Q9F6D6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Acetyl carboxylase.
GN ZHUF.
OS Streptomyces sp. R1128.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=140437;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1128;
RC MEDLINE=20517899; PubMed=10931852;
RA Marti T., Hu Z., Pohl N.L., Shah A.N., Khosla C.;
RT "Cloning, Nucleotide Sequence, and Heterologous Expression of the
RT Biosynthetic Gene Cluster for R1128, a Non-steroidal Estrogen Receptor
RT Antagonist. Insights Into an Unusual Priming Mechanism."
RL J. Biol. Chem. 275:33443-33448(2000).
DR EMBL; AF293442; AAC30193.1; -
DR InterPro; IPR001095; Ac-CoA carboxylA.
DR InterPro; IPR000438; ACOACC_transf.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF03255; ACCA; 1.
DR Pfam; PF01039; Carboxyl_trans; 1.
DR PRINTS; PR01069; ACCCTFRASEA.
DR PRINTS; PR01070; ACCCTFRASEB.
DR TRFPAWS; TTR00515; accd; 1.
SQ SEQUENCE 578 AA; 60667 MW; 979330136912C323 CRC64;
```

Query Match 24.5%; Score 599.5; DB 2; Length 578;
Best Local Similarity 34.2%; Pred. No. 7.9e-28;
Matches 175; Conservative 71; Mismatches 218; Indels 47; Gaps 9;

```
QY 21 IDSVLPDPSFISMNTPQYDNLNOGYAET-----LERARAKACDESITGEGTVEGI 73
D 21 IDSVLPDPSFISMNTPQYDNLNOGYAET-----LERARAKACDESITGEGTVEGI 73
QY 52 LDQLLDPSGSAEPRIINDVTRFDVL--GFTDTPYPARLAARAATGLDEAVIVTGRIGHQ 109
D 52 LDQLLDPSGSAEPRIINDVTRFDVL--GFTDTPYPARLAARAATGLDEAVIVTGRIGHQ 109
QY 74 PVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMMV 133
D 74 PVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMMV 133
QY 110 PLVVAAMDPAFLGSLGAGAVGELLTTLAETALAEHLPLLVASGARMOEGAVVALQMA 169
D 110 PLVVAAMDPAFLGSLGAGAVGELLTTLAETALAEHLPLLVASGARMOEGAVVALQMA 169
QY 134 SITAAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELTGG 193
D 134 SITAAVORHREAHLPVLYLRNPTMGAMASWGSSGHLTFAPGAQIGFLGRVVELTGG 193
```

```

Db      170 KTSALGQDEAGVLTISLVTDPTYGVVAAAFATLCLDVABPGARLGFAGRRVIAQTIR 229
      194 HALPGVOQAEMLVKTGVIGVSPLOLEAAVAKTKV-----IQPEATDRFSPT 244
      220 BELPPEFGTAELHSHGLIDLIVPRRLDELARLRACAKPKGGDGGPARGGLDAGA 289
      245 TPGVALPV-----MEAIARSRDPQRPFGIGIMETLGVKLSGARAGALSPAVRVL 297
      290 SPYVVDPADLAHNDRSQVRAARAIIDRPTLDVIALCLDFELKDRISGDCPAIVAGL 349
      298 ARIGRPVVLIGOD-----RRFTLG---PQELRPARGISLARELMLPVSIIIDTSG 346
      350 ARFGGRPAVAVIGHQKHDAQELRRRNFQMPISGYRKARIMRLADKLGLPVTFVDTPTG 409
      347 AELSGAAEELGIASSIARTLSKLDIAPLPTVSIIIGCGVGGALMLPADLYAANML 406
      410 AYPGAVAEEQGAVALAENLRIMAGLRVAVTVTGGSGGALALAAVAVLMEENGTY 469
      407 SALPEGASAILFRPTNHAELIERQGVQAHLLSGGLDGIVAETE-----HFVEE 458
      470 SVISEGCAILIMQPPAAAGRAARALRYARRELLGIGIVDVLPPEGGTGADPLRADA 529
      459 ILGTISNALISE---LDNNPERAGDSRPTRF 486
      530 LRDAIAGLAELSLPDGHQVLVADRWARFRRY 560

```

RESULT 4

Q9LOAI PRELIMINARY; PRT; 458 AA.

```

AC Q9LOAI;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DN Putative acetyl CoA carboxylase (Alpha and beta subunits).
OS SC02445 OR SCC24.16.
OS Streptomyces coelicolor.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
NCBI_TaxID=1902;
RN NCBI
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Jablonowski E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Metzorroek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;

```

```

RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RT Nature 417:141-147(2002).
DR EMBL; AL939112; CAB86110.1; -.
DR InterPro; IPR000438; AC0ACC_transfB.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF01039; Carboxyl_1.
DR PRINTS; PR01070; ACCCTRFRASEB.
DR Complete proteome.
SQ SEQUENCE 458 AA; 47180 MW; E5E4086CB29B69C1 CRC64;

```

Query Match 19.6%; Score 480; DB 16; Length 458;
 Best Local Similarity 31.1%; Pred. No. 7.9e-21;
 Matches 138; Conservative 57; Mismatches 199; Indels 56; Gaps 9;

```

      16 SALTLIDSLVDPDSITSNETFPQYDNLNOGIAETLERARSKACDESYITGEVTEGIPV 75
      21 SELPHADPWPWBDPGLG-----PGVDASPARAAERTGEASEVVCCTGLVEGTTRA 70
      76 AVILSDPSFLGSGIGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMYSI 135
      71 VLVSTFEGFLGSLGSHRTGDRLEAAYAYAREHRLFPVPLVATGSRMQEMLALTQLORV 130
      136 TAAVQRRHEAHLPELVYLRNFTMGAMASWSSGHLTFAEPGAQIGFLGPRVVELTTGHA 195
      131 ARQSALTRAAAGLAQIAVVRDPATGGGWAITLGAADVVALALGAGVGFAGSRV-----R 183
      196 LPD---GVQAEMLVKTGVITIGIVSPLOLRAAVAKTKV-----QPEATDRF 241
      184 PPDADPAAYTAEQAVAGSADAVVPPGELRATLGRWILATLTPASGPAVPEPLGARD-- 241
      242 SPTTPGVALPV--MEAIARSRDPQRPFGIGIMETLGVKLSGARAGALSP-ARVALA 298
      242 -----LPGMGAEVARARDPGRPRAGAYIDAVFTRVVLSGRCGRDPBGMLCGRG 293
      299 RIGGRPVVLIGODRRFTLGPOELRFRARGISLARELMLPVSIIIDTSGAELSQAAEELGI 358
      294 SHAGRTVAVAQOTGAT- RPAGYRTATLHLADRLGIPVLTVDTPPAANDAEERAGA 352
      359 ASSIARTLSKLDIAPLPTVSIIIGCGVGGALMLPADLYAANMLSALPPEGASAIL 418
      353 GPATADLFGAVASARTPTTVITIGCGSGGALALAPGSTWATPDSYVIAPE----- 406
      419 FRDTHAAELIERQGVQAHLLSQ 442
      407 -----HAAAILKRPPEAVEATAGQ 425

```

RESULT 5

Q8YE58 PRELIMINARY; PRT; 301 AA.

```

AC Q8YE58;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2).
GN BME12020.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
NCBI_TaxID=29459;
RN NCBI
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=1175668;
RA DelVecchio V.G., Kaparatou V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Nazur M., Goltzman E.,
RA Selkov E., Elizer P.H., Hagius S., O'Callaghan D., Teleson J.-D.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
Brucella melitensis.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
 DR EMBL: AEO09634; AAL53201.1; -
 DR InterPro: IPR000438; ACOACC_transfB.
 DR InterPro: IPR000022; Carboxyl_trans.
 DR InterPro: IPR001202; WW_Rsp5_WWP.
 DR Pfam: PF01039; Carboxyl_trans; 1.
 DR PRINTS: PRO1070; ACCCTFRASEB.
 DR TIGRFAMs: TIGR00515; accd; 1.
 DR PROSITE: PS01159; WW_DOMAIN_1; 1.
 KM Ligase; Complete proteome.
 SQ SEQUENCE 301 AA; 33269 MW; BCB69B33AEAD566 CRC64;

Query Match 17.6%; Score 431.5; DB 16; Length 301;
 Best Local Similarity 43.3%; Pred. No. 3.5e-18;
 Matches 88; Conservative 43; Mismatches 69; Indels 3; Gaps 2;

QY 46 YAEFLERARSKAKDESVITGEVIGIPVAVILSDPSFLGSLGTVAASVIMKAIRHAT 105
 DB 93 YIDRLKDYRSRTGMDAIVNGLGTIEGPIVATVQDFSGFMGSLGMGAGEAIIQGFEXAI 152
 QY 106 ELKRLPLVSPASGGARMOEDNRFAVMMVSTIAAQRHREALPLVYLIRNPTMGAMASW 165
 DB 153 ELKRLPLVSPASGGARMOEGILSLMQLPRTTVAEMLEAGLPYIVLTNPTTGGVTASY 212
 QY 166 GSSGHLTFAPGPAQIGFLGRVVELTTHALPDGVQQAENLVKTVIDGIYSPQLRAV 225
 DB 213 AMLGDIHIAEPGALIGFAGRVIEQITREKLPBGFSSEYIMHGMDVMVSRLEKATI 272
 QY 226 AKTLKVI--QPEATDRFSPTTP 246
 DB 273 ARLLKIMTKOPAN--SDAPAPQKP 294

RESULT 6
 Q8FXV7 PRELIMINARY; PRT; 301 AA.

AC Q8FXV7; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Acetyl-CoA carboxylase, carboxyl transferase, beta subunit.
 GN ACED OR BR2107.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seehadi R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.U., Umeyam L., Brinkac L.M., Beaman M.U.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL: AE014498; AAN30997.1; -
 DR TIGR: BR2107; -
 KM Transferrase; Complete proteome.
 SQ SEQUENCE 301 AA; 33269 MW; BCB69B33AEAD566 CRC64;

Query Match 17.6%; Score 431.5; DB 16; Length 301;
 Best Local Similarity 43.3%; Pred. No. 3.5e-18;
 Matches 88; Conservative 43; Mismatches 69; Indels 3; Gaps 2;

QY 46 YAEFLERARSKAKDESVITGEVIGIPVAVILSDPSFLGSLGTVAASVIMKAIRHAT 105
 DB 93 YIDRLKDYRSRTGMDAIVNGLGTIEGPIVATVQDFSGFMGSLGMGAGEAIIQGFEXAI 152
 QY 106 ELKRLPLVSPASGGARMOEDNRFAVMMVSTIAAQRHREALPLVYLIRNPTMGAMASW 165

DB 153 ELKRLPLVSPASGGARMOEGILSLMQLPRTTVAEMLEAGLPYIVLTNPTTGGVTASY 212
 QY 166 GSSGHLTFAPGPAQIGFLGRVVELTTHALPDGVQQAENLVKTVIDGIYSPQLRAV 225
 DB 213 AMLGDIHIAEPGALIGFAGRVIEQITREKLPBGFSSEYIMHGMDVMVSRLEKATI 272
 QY 226 AKTLKVI--QPEATDRFSPTTP 246
 DB 273 ARLLKIMTKOPAN--SDAPAPQKP 294

RESULT 7
 Q9GCN5 PRELIMINARY; PRT; 308 AA.

AC Q9GCN5; 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Acetyl-CoA carboxylase carboxyltransferase beta chain (EC 5.4.1.2).
 GN MRS075.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneke T., Nakamura Y., Sato S., Asamiu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 RL DNA Res. 7:331-338 (2000).
 DR EMBL: AP003005; BAB51586.1; -
 DR InterPro: IPR000438; ACOACC_transfB.
 DR InterPro: IPR000022; Carboxyl_trans.
 DR InterPro: IPR001202; WW_Rsp5_WWP.
 DR Pfam: PF01039; Carboxyl_trans; 1.
 DR PRINTS: PRO1070; ACCCTFRASEB.
 DR TIGRFAMs: TIGR00515; accd; 1.
 DR PROSITE: PS01159; WW_DOMAIN_1; 1.
 DR Transferrase; Ligase; Complete proteome.
 KM Transferrase; Ligase; Complete proteome.
 SQ SEQUENCE 308 AA; 34219 MW; 5FBE6E78E596C500 CRC64;

Query Match 17.4%; Score 427; DB 16; Length 308;
 Best Local Similarity 38.8%; Pred. No. 6.6e-18;
 Matches 93; Conservative 46; Mismatches 87; Indels 14; Gaps 3;

QY 24 VLDPDSFISWNETPQYDNLNOGYAETLERARSKAKDESVITGEVIGIPVAVILSDFS 83
 DB 82 VLDPDKFRD-----EKRYTDRDKDAKATGLEDALVNAGLTIEGPIVATVQDFA 131
 QY 84 FLGSLGTVAASVIMKAIRHATEKLPVLSPPASGGARMOEDNRFAVMMVSTIAAQRH 143
 DB 132 FMGSLTMAAGDAIVHFEVALQRKRLIFASGGARMOEGILSLMQLPRTTGVGRLK 191
 QY 144 EAHLPVLYIRNPTMGAMASWSSGHLTFAPGPAQIGFLGRVVELTTHALPDGVQQA 203
 DB 192 EAGLPYIVLTNPTTGGVTASYALGVHIAERGALIGFAGRVIEQITREKLPDGFORS 251
 QY 204 ENLVKTVIDGIYSPQLRAVAKTLKVIQPEATDRFSPTTGVALPVMKAIRSDPQ 263
 DB 252 EYIMEHGMDVMVSRLEMRQITARLLKMLKMPGEQK--PLEPILPPAV--VAEARPQ 307

RESULT 8
 Q9GQO8 PRELIMINARY; PRT; 1338 AA.

AC Q9GQ08; 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Putative acetyl-CoA carboxylase/pyruvate carboxylase fusion protein.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
 ON NCBI_TaxId=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WB/ATCC9597;
 RT "A putative Acetyl-CoA Carboxylase/Pyruvate Carboxylase fusion protein in the amitochondriate protist, Giardia lamblia."
 RL Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF207740; AAG35712.1; -
 DR InterPro; IPR001095; Ac-CoA carboxylase.
 DR InterPro; IPR000438; ACoACCoA transFB.
 DR InterPro; IPR000089; Biotin lipoyl.
 DR InterPro; IPR000022; Carboxyl trans.
 DR InterPro; IPR000891; HMGCoA-like.
 DR InterPro; IPR003379; PyC-ODA.
 DR Pfam; PF00364; biotin lipoyl; 1.
 DR Pfam; PF01039; Carboxyl trans; 1.
 DR Pfam; PF02436; HMGCoA-like; 1.
 DR PRINTS; PRO1069; ACCTPRPASA.
 DR PRINTS; PRO1070; ACCCTPRPASEB.
 DR PRINTS; PRO1070; ACCCTPRPASEB.
 SQ SEQUENCE 1338 AA; 148221 MW; 8DB0125EA5FAED1 CRC64;

Query Match 16.7%; Score 409; DB 5; Length 1338;
 Best Local Similarity 27.5%; Pred. No. 6.1e-16;
 Matches 132; Conservative 90; Mismatches 192; Indels 66; Gaps 15;

QY 30 FISMNETPOYDNLNQG--YAETLERAR--SKAKDESVITGEGEIPVAVILDSFSL 85
 DB 9 FINLSTHMDVFFVFGVSKQLEAARKTSLCMHYHVEPEYIRA--VQEFAM 65
 QY 86 GGSLSCTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMMVSTAAVORHBA 145
 DB 66 GGSLSCTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMMVSTAAVORHBA 145
 QY 146 HLPFLVLRNPTMGAMASWSSGSHLTPAEFGAIGFLCPRVVELTT-----GHALP 197
 DB 126 GIPVIVLRNPTMGAMASWSSGSHLTPAEFGAIGFLCPRVVELTT-----GHALP 197
 QY 198 DGVOQAEMLVKTGVLDGIVSPLO--LRAAVAKTLKVI-----QPEATD-----RFSPTT 245
 DB 186 AGFQTTDRAAATGALDQLPDSLESAYASIGHILEVLIQKMGQVEAALAKTALHLTGVP 245
 QY 246 PGVALPVMEAI-----ARSRDPORPQIGRIMETLADVVKLSGADAGALSPAVRYAL 297
 DB 246 HGAEDVTEDAFVQAFVYRDRNVTROPQCDYLKMF-----VMAASLARDKICITAYAF 299
 QY 298 ARIGGR-FVVLIGORRFTLQGF-----LRFRRGISLARENLPIVSIITDSGABL 349
 DB 300 GVFAFDTLVIALATSRREDVLPDEAGLSPGTGRFAARAIYLANRLSIPVITLVDPAALP 359
 QY 350 SQAAEELGIASSIARTLSKLIDAPLPVSVIIGGVGGALMLADIVYAENMNLAL 409
 DB 360 SPEAEEOGOSRAISCLAFQAVHYPIISITIGEGSGGALLAGAGVGLISKAFVYVI 419
 QY 410 PPEGASAIL-----FDTNHAET-----IEROGVQVNA-----LLSQGLIDGIVAEET 454
 DB 420 SPEGASAILQPSVYR--NNTAEMRANFISDAELLAHVQRCVPIRNNGVVNDIIVGEY 477

RESULT 9
 Q8GDQ2 PRELIMINARY; PRT; 272 AA.
 AC Q8GDQ2;
 AC Q8GDQ2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (EC 6.4.1.2) (Fragment).
 OS Hellobacillus mobilis.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Hellobacteriaceae;
 ON NCBI_TaxId=28064;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22337798; PubMed=12446909;
 RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
 RA Blankenship R.E.;
 RT "Whole-genome analysis of photosynthetic prokaryotes."
 RL Science 298:1616-1620 (2002).
 DR EMBL; AY142924; AAN87528.1; -
 DR Transferase; Ligase.
 KW NON TER 272
 FT 272
 SQ SEQUENCE 272 AA; 29796 MW; 721D9DA2EA18812D CRC64;

Query Match 16.5%; Score 404; DB 2; Length 272;
 Best Local Similarity 41.5%; Pred. No. 1.3e-16;
 Matches 90; Conservative 40; Mismatches 85; Indels 2; Gaps 2;
 QY 21 IDSVLDPDSFISWN-ETPOYDNLN-QGYAETLERARSKAKDESVITGEGEIPVAVI 78
 DB 50 IDSLIDEGFTIEMDELVTQDPLNPPGYAQKIQKQOSTMAEALITQGGIHLIPVAVI 109
 QY 79 LSDSFLLSGLTAVSVRIMKAIHRATELKLPLVSPASGARMOEDNRAFVMMVSTAA 138
 DB 110 AMDSRFIWASGAVVGEKIVAAEKALEKMPILIFSASGARMOEGVLSLMQARTSAA 169
 QY 139 VQHRREAHLPVLRNPTMGAMASWSSGSHLTPAEFGAIGFLCPRVVELTTGHALPD 198
 DB 170 LTRLEAGLPFVSVLTPTPTGTVASFMALDILLIIEAGALIGFTGPRVIRQTLRKLPA 229
 QY 199 GVOQAEMLVKTGVLDGIVSPLO--LRAAVAKTLKVIQV 235
 DB 230 GFQSEFLQKHGWVDIIEPRKMAQLATLLSLHMPM 266

RESULT 10

QY 08YU12 PRELIMINARY; PRT; 316 AA.
 AC 08YU12;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Acetyl-CoA carboxylase beta subunit.
 GN ACCD OR ALU2364.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 ON NCBI_TaxId=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21595285; PubMed=11759840;
 RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120."
 RL DNA Res. 8:205-213 (2001).
 DR EMBL; AP003589; BAB74063.1; -
 DR InterPro; IPR000438; ACoACCoA transFB.
 DR InterPro; IPR000022; Carboxyl trans.

DR Pfam; PF01039; Carboxyl trans; 1.
 DR PRINTS; PRO1070; ACCCTFRASEB.
 DR TRIGFAMS; TIGR00515; accd; 1.
 KW Complete proteome.
 SQ SEQUENCE 316 AA; 35455 MW; FDB192P20BDF633E CRC64;

Query Match 16.0%; Score 392.5; DB 16; Length 316;
 Best Local Similarity 38.0%; Pred. No. 7.9e-16;
 Matches 87; Conservative 42; Mismatches 81; Indels 19; Gaps 3;

OY 32 SWNEPTQDNL-----NQYAETLERARSAKCDSEYITGEVEGIPVAVILSD 81
 DB 85 TW--PMENENRATPDLPQFRDKAVSDRLREMEDTGLIDAVKTIAGQINSPVALAVMD 142
 OY 82 FSLFGSLGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAFVMMVSIITAAVOR 141
 DB 143 RFPMGSGSVGEEKITRIIEATQRRIPVVICTSGARMQEGMLMOMAKISALIER 202
 OY 142 HREAHPLPLVLYLRNPTMGAMASWGSGLTFPAEPGAQIGFLGPRVVELTTGHALPDGVQ 201
 DB 203 HRDARLVIPLVLTNPTTGVTSFAMLDIILAEFKATIGFAGRVITQTLREKLPDDPQ 262
 OY 202 QAEINVKTGVIDIGVSPLOLRVAVKTKVIOPEATDRFSPTPGVAL 250
 DB 263 TAEDLKHGFVDIVPRITQLKNITLSQLIALHQV-----FTTPPMVL 304

RESULT 11

OY 092TC7 PRELIMINARY; PRT; 304 AA.
 AC 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Probable acetyl-coenzyme A carboxylase carboxyl transferase subunit
 GN beta protein (EC 6.4.1.2).
 OS ACCD OR R00030 OR SMC02764.
 OS Rhizobium melliceti (Sinorhizobium melliceti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OK NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Batloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
 Pohl T., Portetleau D., Puehler A., Purnelle B., Ramsperger U.,
 Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium melliceti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
 DR EMBL; AL591782; CAC41417.1; -
 DR InterPro; IPR000438; ACOACC.transfB.
 DR InterPro; IPR000022; Carboxyl.trans.
 DR Pfam; PF01039; Carboxyl.trans.1.
 DR PRINTS; PRO1070; ACCCTFRASEB.
 DR TRIGFAMS; TIGR00515; accd; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 304 AA; 33567 MW; 4FCC45B674D068AB CRC64;

Query Match 15.9%; Score 389.5; DB 16; Length 304;
 Best Local Similarity 39.2%; Pred. No. 1.1e-15;
 Matches 82; Conservative 44; Mismatches 80; Indels 3; Gaps 2;

OY 46 YAEETLRARSAKCDSEYITGEVEGIPVAVILSDPFLGSLGTVAVRIMKAIHRAT 105
 DB 93 YSDRARDRAKTELEDTIVAGIGQIKLVAVAHENFPGSLIAGAEIVAKFEPAI 152
 OY 106 ELKPLPLVSPASGARMQEDNRAFVMMVSIITAAVORHREAHPLPLVLYLRNPTMGAMASW 165
 DB 153 AEKPLVMPASGARMQEGILSLQLPRTVALNMLKEAGIPYIVLTNPTTGVTSY 212

OY 166 GSSGHLTFEPGAQIGFLGPRVVELTTGHALPDGVQOAEINVKTGVIDIGVSPLOLRVAV 225
 DB 213 AMLGDIMAEBAIGAGVIGAGRVIRIQTIREKLPESQFSTILEHGMVDMVVKRHDIEFTL 272
 OY 226 AKTLKVI--OPEATDRFSPTPGVALPV 252
 DB 273 ARVINIIMKKPAKAVKR-DTATELAPLPV 300

RESULT 12

OY 092KQ1 PRELIMINARY; PRT; 289 AA.
 AC 092KQ1
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Acetyl-coenzyme A carboxylase subunit B.
 GN ACCD OR JHP0884.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OK NCBI_TaxID=85963;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 Smith D.R., Noonan B., Guild B.C., deJonge B.U., Carmel G.,
 Tummino P.J., Caruso A., Urie-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 Trust T.J.,
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori."
 RL Nature 397:176-180 (1999).
 DR EMBL; AE001518; AAD06468.1; -
 DR InterPro; IPR000438; ACOACC.transfB.
 DR InterPro; IPR000022; Carboxyl.trans.
 DR Pfam; PF01039; Carboxyl.trans.1.
 DR PRINTS; PRO1070; ACCCTFRASEB.
 DR TRIGFAMS; TIGR00515; accd; 1.
 KW Complete proteome.
 SQ SEQUENCE 289 AA; 32567 MW; 43B6EAB878E630 CRC64;

Query Match 15.8%; Score 387.5; DB 16; Length 289;
 Best Local Similarity 39.6%; Pred. No. 1.4e-15;
 Matches 91; Conservative 43; Mismatches 81; Indels 15; Gaps 4;

OY 16 SALTLLIDVLDPPSFLISMN--TPQYDNLN-----QGYAETLERARSAKCDSEYITGE 69
 DB 60 NATRIEFLDVGSEFEFDKHLRN-DPLNFVDKESYKQIKKYEKKTNPSSVIGSEAK 118
 OY 70 VEGIPVAVILSDPFLGSLGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAF 129
 DB 119 INRMPLQIVVDFEFMGSGSVGEKIVRAINRAVAKKALLIVSASGARMOESTYSL 178
 OY 130 VMVVSITAAVORHREAHPLPLVLYLRNPTMGAMASWGSGLTFPAEPGAQIGFLGPRVVE 189
 DB 179 MOWAKTSBALNRLSEALKPFLISLSDPTYGVSASPFLEGLIILAECAATIGFAGRVIK 238
 OY 190 LTTGHALPDGVQOAEINVKTGVIDIGVSPLOLRVAV-----KTLKV 231
 DB 239 QTIQADLPFGQTAEPFLHEGLIDMIVHRDKDKTISDLIAMWTHKTSKI 288

RESULT 13

OY 08DIE7 PRELIMINARY; PRT; 309 AA.
 AC 08DIE7
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Acetyl-CoA carboxylase beta subunit.
 GN ACCD OR TLR1643.

OS *Synechococcus elongatus* (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 NCBI_TaxId=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 Watanabe A., Iriuchi M., Kawashima K., Kimura T., Kishida Y.,
 Kiyokawa C., Kohara M., Matsumoto M., Matsumo A., Nakazaki N.,
 Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1."
 RL DNA Res. 9:123-130(2002).
 DR EMBL, AP005374; BAC09195.1; -
 SQ Complete proteome.
 Query Match 15.8%; Score 386; DB 16; Length 309;
 Best Local Similarity 37.0%; Pred. No. 1.9e-15;
 Matches 94; Conservative 46; Mismatches 96; Indels 18; Gaps 5;
 QY 11 GHEH-TSALTLDVIDPDSFISWNETPQYDNI-----NCGYATLEBRASKAC 59
 DB 56 GHHHRVSSSERIQQLIDPH---TWR--PDENLVSCDPIQFKDKRKYSDRLREYQKGTGL 110
 QY 60 DESVTIGEGTVRGIPYAVILSDPSFISWNETPQYDNI-----NCGYATLEBRASKAC 119
 DB 111 KQAVQTGLQLELPALGMDPAFVGSGMGSVGEKITRLIERAWEHIPLIVCASGG 170
 QY 120 AEMOEDNRAFYVMVNSITAAVORHREALPLVLYRNPTMGAMASGSSGHLTPAEPGAQ 179
 DB 171 AAMQSGMLSLMGMAKRAALERRHRSAGLYIPILHTPTAGYASAMGDIITIAEPKAT 230
 QY 180 IGFLEPRVELTTGHALPDGVOAENLVKTVGIDGIVSPQLRAAVAKTLKYIOPVEARD 239
 DB 231 IAFAGRRVIEQTLREKLPDFAEFVQKCGFVDIVPRTQKSTLARILRLHQVPS-- 288
 QY 240 RPSPTTPGVALPVM 253
 DB 289 RMSALLPKSFPLM 302

RESULT 14
 025604 PRELIMINARY; PRT; 289 AA.
 AC 025604;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Acetyl-CoA carboxylase beta subunit (ACCD).
 GN HP0950.
 OS *Helicobacter pylori* (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 Pletschmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
 Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,
 Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 pylori".
 RL Nature 388:539-547(1997).
 DR EMBL, AE00604; AAD07995.1; -

DR TIGR, HP0950; -
 DR InterPro; IPR000438; ACOACC_transf.
 DR InterPro; IPR000022; Carboxyl_trans.
 DR Pfam; PF01039; Carboxyl_trans; 1.
 DR PRINTS; PRO1070; ACCCTRRPASEB.
 DR TIGRPFams; TIGR00515; accd; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 289 AA; 32553 MW; A57ADDAF3AB0AFC CRC64;
 Query Match 15.7%; Score 384; DB 16; Length 289;
 Best Local Similarity 39.3%; Pred. No. 2.3e-15;
 Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;
 QY 26 DDPSFISWNETPQYONLNGYVETLEBRASKACDESVTIGEGTVGIPYAVILSDPSFL 85
 DB 85 DPLNFDV-----KSYKQRIKYEKRTNRPSSVSGAKINRMLQIVDFDFSPW 134
 QY 86 GSGTGVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFYVMVNSITAAVORHREA 145
 DB 135 GSGTGVSEGEKIVRAINAVAREALLIVSAGGARMQESTVSLQMATSAALRLSEA 194
 QY 146 HLPFLVLYRNPTMGAMASGSSGHLTPAEPGAQIGFLEPRVELTTGHALPDGVOAEN 205
 DB 195 KLPFTSLSDPFGVGSASFALGDLITAEFGAMIGFAGPRVYKOTIGADLPFGFOTAEF 254
 QY 206 LKTVGIDGIVSPQLRAAVA-----KTLKY 231
 DB 255 LIEHGLIDVTVARKDKTSLDILMMTHTKSKI 288

RESULT 15
 050132 PRELIMINARY; PRT; 201 AA.
 AC 050132;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE znfa.
 OS *Mycobacterium leprae*.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxId=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith D.R.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 DR EMBL; U15184; AAA63045.1; -
 DR InterPro; IPR000438; ACOACC_transf.
 DR InterPro; IPR000022; Carboxyl_trans.
 DR Pfam; PF01039; Carboxyl_trans; 1.
 DR PRINTS; PRO1070; ACCCTRRPASEB.
 SQ SEQUENCE 201 AA; 21323 MW; EF45B8DCBB804FB4 CRC64;
 Query Match 15.5%; Score 379; DB 2; Length 201;
 Best Local Similarity 44.4%; Pred. No. 2.7e-15;
 Matches 87; Conservative 28; Mismatches 73; Indels 8; Gaps 3;
 QY 122 MOEDNRAFYVMVNSITAAVORHREALPLVLYRNPTMGAMASGSSGHLTPAEPGAQIG 181
 DB 1 MOEGTVAFLQWVKIAAAVAKLSKRAGLPYLGYLRNPTTGGVPSWSLGHVTVAEFGVLIG 60
 QY 182 FLEPRVELTTGHALPDGVOAENLVKTVGIDGIVSPQLRAAVAKTLKYIOPVEATDRF 241
 DB 61 FLGPRVVELLYGEPSPDIOGTANENLQRGIVDAIVTLDGLQTLNRLATMADV---PKL 117
 QY 242 SPT---TPGVALPVEAIARSRDPQRPDGEIMETTGADVVKLSGARAGALSPAVRYAL 297
 DB 118 IPTQRPPEIPDVDAWNSVMSRRPSPVQAVLRHGATDVLVLSGPGHGAATTL-LAL 176

QY	298	ARIGRPVLLIGDDR	313
		: : : :	
Db	177	ARLAGTAVVIGQQRK	192

Search completed: November 13, 2003, 06:42:46
Job time : 94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:41:05 / Search time 29 Seconds
(without alignments)
716.367 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKRPPTVMGMEHTSALTLL.....NNPERAGDRSFRFERRLAQ 491

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTCUTS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2448	100.0	491	US-09-362-899-3	Sequence 3, Appl1
2	413	16.9	215	US-08-311-731A-410	Sequence 410, App
3	354.5	14.5	293	US-09-134-001C-3803	Sequence 3803, Ap
4	354	14.5	289	US-09-107-532A-3824	Sequence 3824, Ap
5	331	13.5	300	US-09-328-352-5240	Sequence 5240, Ap
6	328.5	13.5	339	US-09-252-991A-28751	Sequence 28751, A
7	319.5	13.1	255	US-09-198-452A-76	Sequence 76, Appl1
8	289	11.8	317	US-08-973-375-1	Sequence 1, Appl1
9	286	11.7	279	US-09-328-352-8167	Sequence 8167, Ap
10	283.5	11.6	315	US-09-134-001C-3786	Sequence 3786, Ap
11	271.5	11.1	262	US-09-107-532A-4837	Sequence 4837, Ap
12	265.5	10.8	344	US-09-198-452A-433	Sequence 433, App
13	262.5	10.3	327	US-08-422-560A-12	Sequence 12, Appl1
14	252.5	10.3	327	US-08-468-793-12	Sequence 12, Appl1
15	252.5	10.3	327	US-08-468-793-12	Sequence 12, Appl1
16	250	10.2	332	US-09-252-991A-21222	Sequence 21222, A
17	215.5	8.8	543	US-08-776-597A-2	Sequence 2, Appl1
18	215.5	8.8	543	US-08-693-228-2	Sequence 2, Appl1
19	210	8.6	546	US-09-328-352-5584	Sequence 5584, Ap
20	198	8.1	544	US-09-252-991A-26096	Sequence 26096, A
21	192	7.8	1116	US-09-252-991A-24374	Sequence 24374, A
22	186.5	7.6	78	US-08-311-731A-371	Sequence 371, App
23	178	7.3	692	US-09-252-991A-27997	Sequence 27997, A
24	177.5	7.3	538	US-09-328-352-6530	Sequence 6530, Ap
25	176.5	7.2	566	US-09-107-532A-5786	Sequence 5786, Ap
26	155.5	6.4	332	US-09-328-352-6531	Sequence 6531, Ap
27	144	5.9	336	US-09-252-991A-16707	Sequence 16707, A

28	143.5	5.9	5215	3	US-09-105-537-2	Sequence 2, Appl1
29	128.5	5.2	391	4	US-09-252-991A-24230	Sequence 24230, A
30	126.5	5.2	2254	2	US-08-677-010-3	Sequence 3, Appl1
31	126.5	5.2	2254	2	US-08-790-519-3	Sequence 3, Appl1
32	122	5.0	51	1	US-08-418-893D-4	Sequence 4, Appl1
33	122	5.0	120	4	US-08-311-731A-411	Sequence 411, App
34	122	5.0	3201	4	US-09-679-279-15	Sequence 15, Appl1
35	121	4.9	3739	3	US-09-320-878-2	Sequence 2, Appl1
36	121	4.9	3739	3	US-09-105-537-33	Sequence 33, Appl1
37	121	4.9	3739	4	US-09-141-908-3	Sequence 3, Appl1
38	121	4.9	3739	4	US-09-657-440-2	Sequence 2, Appl1
39	121	4.9	11877	3	US-09-105-537-6	Sequence 6, Appl1
40	119.5	4.9	841	3	US-09-413-814-107	Sequence 107, App
41	119.5	4.9	4545	2	US-08-804-227C-14	Sequence 14, Appl1
42	119.5	4.9	4550	2	US-08-804-227C-8	Sequence 8, Appl1
43	119.5	4.9	4550	2	US-08-804-198-2	Sequence 2, Appl1
44	118	4.8	822	4	US-09-252-991A-23090	Sequence 23090, A
45	118	4.8	2628	3	US-09-413-814-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1									
US-09-362-899-3									
Sequence 3, Application US/09362899									
Patent No. 6361986									
GENERAL INFORMATION:									
APPLICANT: Degussa-Hls AG									
APPLICANT: Forschungszentrum-Jilich GmbH									
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND									
CURRENT APPLICATION NUMBER: US/09362,899									
EARLIER FILING DATE: 1999-07-29									
EARLIER APPLICATION NUMBER: DE 19924365.4									
NUMBER OF SEQ ID NOS: 3									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 3									
LENGTH: 491									
TYPE: PRT									
ORGANISM: Corynebacterium glutamicum									
US-09-362-899-3									
Query Match									
Best Local Similarity 100.0%; Pred. No. 3.5e-242; Length 491;									
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	VEKRPPTVMGMEHTSALTLLIDSVDPSFTSMNETPOYDNLNOGYAETLEBRASKACD	60						
DB	1	VEKRPPTVMGMEHTSALTLLIDSVDPSFTSMNETPOYDNLNOGYAETLEBRASKACD	60						
QY	61	ESVITGEGTVEGIPAVILSDPSFLGSLGTVAIVRINKAIHRATELKLPLVSPASGA	120						
DB	61	ESVITGEGTVEGIPAVILSDPSFLGSLGTVAIVRINKAIHRATELKLPLVSPASGA	120						
QY	121	RMQENRFAVMNVSTTAAVQHRREHLPLVYLNRPTMGMASSGSGHITPAEPGQI	180						
DB	121	RMQENRFAVMNVSTTAAVQHRREHLPLVYLNRPTMGMASSGSGHITPAEPGQI	180						
QY	181	GFLGRVVELTGHALPDGVOAENLVYTGVIDGIVSPLOQRAAVAKTLKYOPEATDR	240						
DB	181	GFLGRVVELTGHALPDGVOAENLVYTGVIDGIVSPLOQRAAVAKTLKYOPEATDR	240						
QY	241	FSPPTPGVALPVEAIAASRDPORGIGETIMETIGADVYKLSGARAGALSPAVVALARI	300						
DB	241	FSPPTPGVALPVEAIAASRDPORGIGETIMETIGADVYKLSGARAGALSPAVVALARI	300						
QY	301	GGRPVVILGDRRRTLGGQELRPARGISLARETNPVSIIDPSGALSGAABELGIAS	360						
DB	301	GGRPVVILGDRRRTLGGQELRPARGISLARETNPVSIIDPSGALSGAABELGIAS	360						
QY	361	STARTSLIDAPAPPTVSVIIGQGVGGALMLPADVYVAENMWLSLPEGASATIFR	420						
DB	361	STARTSLIDAPAPPTVSVIIGQGVGGALMLPADVYVAENMWLSLPEGASATIFR	420						

Db 361 STARTLSKLIDAPLFTSVIIGGVGGALAMLPA DLVYAAENAMLSALPEGASAILFR 420
Qy 421 DPNHAEIIRERGQAHALLSGGLDGIYAETHEVEELGTISNAISLDNNPERAGD 480
Db 421 DPNHAEIIRERGQAHALLSGGLDGIYAETHEVEELGTISNAISLDNNPERAGD 480
Qy 481 SRFTFRERLAQ 491
Db 481 SRFTFRERLAQ 491

RESULT 2
US-08-311-731A-410
; Sequence 410, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 410:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae
; US-08-311-731A-410

Query Match 16.9%; Score 413; DB 4; Length 215;
Best Local Similarity 45.6%; Pred. No. 3.9e-34;
Matches 94; Conservative 29; Mismatches 75; Indels 8; Gaps 3;

Qy 112 LVSPASGARMQEDRAFMVMSITAAVQRRREALPLVLYLRNPTMGAMASGSGHL 171
Db 5 LVSPSSDSTRMQEGVALFLQWKIAAAVKLSKRAGLPYLGILRNPTTGGVSPASGSLGHV 64
Qy 172 TPAEBGAIQIGLPFVVELTGGHALPDGVQAENLVKTGVIDGIVSPQLRAAVAKTLKV 231
Db 65 TVABEGVILGIFLPVVELLYGEPSPSDIQTAENLQHRGVIDATVLDGLQLTLNRALTM 124
Qy 232 IQPHEATDRFSPPT---TPGVALPMEALIASRDPQRCIGIEIMETTLGADVYKLSGARAG 287
Db 125 IADV---PKLPTQRPSPIDPVPAWMSVMSRRERPSVAQVLRHGATDVLLSGPHG 181
Qy 288 ALSPAVRVALARIGGRPVVLIGQDR 313

Db 182 EAATTL-LALARLAGQTAIVYIGQGR 206

RESULT 3
US-09-134-001C-3803
; Sequence 3803, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3803
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3803

Query Match 14.5%; Score 354.5; DB 4; Length 293;
Best Local Similarity 36.0%; Pred. No. 6.4e-28;
Matches 82; Conservative 48; Mismatches 93; Indels 5; Gaps 3;

Qy 13 EHTSALTL---IDSLDPDSFISWNETPOYDN-LN-OGVAETLERARAKADESVITGE 67
Db 64 DHIHALTAHYKRIEASISBGSFIERDRGMTSANPLDPFGYEKEIKQOQKGLAEALVSGT 123
Qy 68 GTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIRHATELKLPLVSPASGARMQEDNR 127
Db 124 AKLDGIQYVAVMDARFRMGSMGVSVEKICRIIDYCTEHRPLPILSPASGARMQEGII 183
Qy 128 AFVMMVSTIAVQRRREALPLVLYLRNPTMGAMASGSGHLTPAEBGAIQIGLPFV 187
Db 184 SLMGKTSVSLKSHSPAGLGLYSITNPPTTGGVSPASGSLGHVSEPRALIGFAGRRV 243
Qy 188 VELTGGHALPDGVQAENLVKTGVIDGIVSPQLRAAVAKTLKVIOPV 235
Db 244 IEQTINKEPLPDDFTAFLEHGLDKVIRKDMRETLISNLIKHOEV 291

RESULT 4
US-09-107-532A-3824
; Sequence 3824, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match
Best Local Similarity 35.5%; Score 354; DB 4; Length 289;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISWNET-----POYDNLNGYAEFLTERARSKACDESVITGEGTVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTERIKKMDGTGHEAVLTGKATIGIPE 122
QY 75 VAVILSDPSFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGEKITRLPERATKESLPVVLFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLNRPTMGAMASWGSGHLTPFAEPGAGIIGFPRVVELTTGHH 194
DB 183 ISAAVKRNSNEGLFLVLTPTTGGVTSFMSBGDIIIAEPQSILGPRRVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPQLRAAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKIHLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match
Best Local Similarity 13.5%; Score 331; DB 4; Length 300;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSFISWNETQYDNLNGYAEFLTERARSKACDESVITGEGTVEGIPVAVILSDF 82
DB 97 SAKPDLKQVD-----SKYPDRMRKRAQTKGTETALTMQGNINGVDMACAEFF 146
QY 83 SFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVQRH 142
```

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match
Best Local Similarity 35.5%; Score 354; DB 4; Length 289;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISWNET-----POYDNLNGYAEFLTERARSKACDESVITGEGTVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTERIKKMDGTGHEAVLTGKATIGIPE 122
QY 75 VAVILSDPSFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGEKITRLPERATKESLPVVLFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLNRPTMGAMASWGSGHLTPFAEPGAGIIGFPRVVELTTGHH 194
DB 183 ISAAVKRNSNEGLFLVLTPTTGGVTSFMSBGDIIIAEPQSILGPRRVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPQLRAAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKIHLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match
Best Local Similarity 13.5%; Score 331; DB 4; Length 300;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSFISWNETQYDNLNGYAEFLTERARSKACDESVITGEGTVEGIPVAVILSDF 82
DB 97 SAKPDLKQVD-----SKYPDRMRKRAQTKGTETALTMQGNINGVDMACAEFF 146
QY 83 SFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVQRH 142

; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match
Best Local Similarity 35.5%; Score 354; DB 4; Length 289;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISWNET-----POYDNLNGYAEFLTERARSKACDESVITGEGTVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTERIKKMDGTGHEAVLTGKATIGIPE 122
QY 75 VAVILSDPSFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGEKITRLPERATKESLPVVLFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLNRPTMGAMASWGSGHLTPFAEPGAGIIGFPRVVELTTGHH 194
DB 183 ISAAVKRNSNEGLFLVLTPTTGGVTSFMSBGDIIIAEPQSILGPRRVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPQLRAAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKIHLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match
Best Local Similarity 13.5%; Score 331; DB 4; Length 300;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSFISWNETQYDNLNGYAEFLTERARSKACDESVITGEGTVEGIPVAVILSDF 82
DB 97 SAKPDLKQVD-----SKYPDRMRKRAQTKGTETALTMQGNINGVDMACAEFF 146
QY 83 SFLGSLGCTVAVIRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVQRH 142
```

```

; NAME/KEY: SITE
; LOCATION: 1...255
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-76

```

Query Match	13.1%	Score 319.5	DB 4	Length 255
Best Local Similarity	38.7%	Pred. NO. 2e-24		
Matches 75, Conservative	30	Mismatches 74	Indels 15	Gaps 3

Qy	16	SAATLLDSDVADPDSFISMETPOYDUL-----NCGAETETARRASFAKDESEVIT	65
Db	60	TAIERVATLADKO---SMR--PLYTDKSGQDPLEFIDIDTYANRREKARKNTTESSEVIV	114
Qy	66	GEGETBEGIPYAVATLSDPSFLGSGLGTVASVIRIMKAIIRATEIKULPLVSPASGARGNOED	125
Db	115	GICTIGHPAALAVMPDENFMAGSGMAVYCKLTRLEEIEIRLPLVLIISASGARGMOES	174
Qy	126	NRAFWMMVSTTAAVQGHRENAHLPLVLYLNLNPTMGAMMSKSSGHLTFAEFGAQTGLGP	185
Db	175	VFSIMQWVWKTSAALAKLHEAGLPYISVLNTPISGGTVASFAALGDIIIAERKALICFAGR	234
Qy	186	RVVELTTGHALPDG	199
Db	235	RVVAQVITGEDLPFG	248

RESULT 8

US-08-973-275-1
Sequence 1, Application US/089732275B
Patent No. 5958706
GENERAL INFORMATION:
APPLICANT: MATSUNAGA, Tadashi
APPLICANT: KAMIYA, Shunji
TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS
TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE
TITLE OF INVENTION: THEREOF
FILE REFERENCE: MATSUNAGA
CURRENT APPLICATION NUMBER: US/08/973,275B
CURRENT FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: PCT/Jp97/01043
EARLIER FILING DATE: 1997-03-27
EARLIER APPLICATION NUMBER: JP 8-97536
EARLIER FILING DATE: 1996-03-28
EARLIER APPLICATION NUMBER: JP 8-146833
EARLIER FILING DATE: 1996-05-16
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 317
TYPE: PRT
ORGANISM: Magnetospirillum AMB-1
US-08-973-275-1

Query Match	11.8%;	Score 289;	DB 21;	Length 317;
Best Local Similarity	29.2%;	Pred. No. 3.8e-21;		
Matches 100;	Conservative 56;	Mismatches 130;	Indels 56;	Gaps 10

Oy	17	LTPEPAQIGFGRPVLTGTHALPDG--VOQENLVKTVGIDGIVSPQLQAAVAKT	228
Db	4	LEPEKPLAE--LEKGIIEEL--RHLSDDGDVNIADVEVK-----LDQKDKL	45
Oy	229	LKVLPVEATDRSPPTPGVALPVEAIAIARSRDPORPGIEMETLGADVKLGGARAGA	288
Db	46	LR-----STYAKLTPWQKQV-----AHRKRPHTLAVISPLIEDPFLMGDRFAA	91
Oy	289	LSPRVVALARIIGRPVVLIGDPR-----RFTLG--POLPLFARRGISLARELNP	333
Db	92	EDEAIIIGGLRFRPAASVMIIGHEKGDHDTETRLKHNFGMAKPRGYAKAKLMEADHRCVP	151
Oy	338	IVSIITDSGAELOAAEELGIASISARTLSKLDAPLPVSVITIGGVGGAALMLRPDL	397
Db	152	ITLTVDPAAGVPGVDLEARGOAAIKARSIETCTIANRVPVLSVITIGEGSGGAILATGNT	211

```

OY      398  VYAAENAMTSLALPEBESALILFRDNNHAAEIIERGGVOAHNLSQGLIDGIVAE-----TE 453
Db      212  VLMLEHAIIVSVSPBCASAILMRSAENKKDAEQLRLTMOULHKLSTIISDVPEPPMGAAH 271
OY      454  HEVVEILGTTISNN-----LSELNNNEPRACGRDSRFRFRFR 488
Db      272  RNDPLMMQTSMLDLSALRDLSSGLGGGVLRARRRRREKFLPMGR 313

```

RESULT 9

```

US-09-328-352-8167
; Sequence 8167, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8167
; LENGTH: 279
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8167

```

Query Match	11.7%;	Score 286;	DB 4;	Length 279;
Best Local Similarity	31.3%;	Pred. No. 6.2e-21;		
Matches	88;	Conservative 47;	Mismatches 106;	Indels 40; Gaps 8;

```

OY      232  IOGV-EXTORFSPTTGGVALLPWEALARSPOBORG-----IGETMETLIGADVYLLSARA 286
           | : : : : :
Db      4  IRPWKATOSKATTVYOLA-----KHBPPOFLIDVYGLFTEFDA-----LHSDRL 49
           | : : : : :
OY      287  GALSPAVVALARIGGRPVYLLIGODR-RFT-----LGPOLRPARRGISLARELN 335
           | : : : : :
Db      50  FGDDGAMVGGALRFDDQPMVYVIGQRGRSTREKTKHNGCMNPEGYRKSORLLDMAERFN 109
           | : : : : :
OY      336  LPVYSIITDPSGAELSQAAEEHGIASISARTSKLIDALPVPVSYIIDQVGGGALMLPA 395
           | : : : : :
Db      110  LPVFTFDIDMGALPGVGAERQQAALITSLSAQSLSKVPATATVLABSGSGGALGIGVA 169
           | : : : : :
OY      396  DLVYAAENMWLSALPREGASALIFPEDTHAAEIIERQGVQAHALLSOGLLIDGIVAE-----451
           | : : : : :
Db      170  DRYVIMLSHSIYSVISPQSGASILMTKTAERKAQASBALGTDADKQISGIIVEYVDEGEA 229
           | : : : : :
OY      452  ----TEHFVEIIGTISNALSEL-----DNNPERAGDSRFTFR 486
           | : : : : :
Db      230  HLDPERVMQMLKVYKQALDELLPMDAERCEARQRIIMTKF 270

```

RESULT 10
TTC-09-134

```

US09-134-001C-5/be
; Sequence 3786, Application US/09134001C
; Patent No. 6580370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3786
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

```


RESULT 13

US-08-611-107-12

Sequence 12, Application US/08611107

Patent No. 5801233

GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr

TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING

TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES

TITLE OF INVENTION: THEREFOR

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: United States of America

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/611,107

FILING DATE: Concurrently Herewith

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 07/956,700

FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/09340

FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 08/422,560

FILING DATE: 14-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: ARCD:221

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-611-107-12

Query Match 10.3%; Score 252.5; DB 1; Length 327;

Best Local Similarity 25.2%; Pred. No. 2.2e-17;

Matches 86; Conservative 62; Mismatches 138; Indels 55; Gaps 9;

171 LTFAPGAGIGFLGRVVELTGHALPDGVQAENLVKTGVIDIVSPLOLRAAVAKTLK 230
12 LEFEKPLVE--LEERTIQIRT-----LAADNQVD--VSGIQOLEKRAIQLR--- 54
231 VIQPEVATDRSPPTTPGVALPWEAIAASRDPOFGIGIMETLIGADVVKLSGARAGALS 290
55 -----REIFSNLSIPAQRIGV-----ARHPRRSTLDYIQAISDWEIHLHGDRNGSD 101
291 PAVVALARIGRPRVVLIGOD-----RRFTL-GPQELRPARRGISLARLNLPIV 339
102 LALVGVALDGPVFLGHQGRDTONVLRNFGMASFGYRKALRLMEHADRFGMPL 161
340 SIIDTSGAELSGAAEELGIASSIARTLSKLIDAPLPTVSVIIIGGVGGALAMLPADLVY 399
162 TFDITPGAVAGVSAEELGGGGAIVANLRMEFRFSVPLICTVIGGGSGGALGIVGDRLL 221
400 AAENAMLSALPPEGASAILFRDTNHAETIERQGVQAHALLSGGLIDGIVA----- 450

Db 222 MFEHSVYTVASPEKCAIILWRDAGKAAQAAEAALKITARDKQGIIDEITTEPLGASHA 281
451 --ETEHFVEELIGTISNALSELDDNPPRAGDRSFTTFEFL 489
Db 282 PLETAQSLROYL--LRHLKDIALSPAQLREORYQKFRQL 319

RESULT 14

US-08-422-560A-12

Sequence 12, Application US/08422560A

Patent No. 5910626

GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr

TITLE OF INVENTION: ACETYL-COA CARBOXYLASE COMPOSITIONS AND

TITLE OF INVENTION: METHODS FOR USE

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/422,560A

FILING DATE: 14-APR-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/956,700

FILING DATE: 02-OCT-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wilson, Mark B.

REGISTRATION NUMBER: 37,259

REFERENCE/DOCKET NUMBER: ARCD:152/WIM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512-418-3000

TELEFAX: 512-474-7577

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-422-560A-12

Query Match 10.3%; Score 252.5; DB 2; Length 327;

Best Local Similarity 25.2%; Pred. No. 2.2e-17;

Matches 86; Conservative 62; Mismatches 138; Indels 55; Gaps 9;

171 LTFAPGAGIGFLGRVVELTGHALPDGVQAENLVKTGVIDIVSPLOLRAAVAKTLK 230
12 LEFEKPLVE--LEERTIQIRT-----LAADNQVD--VSGIQOLEKRAIQLR--- 54
231 VIQPEVATDRSPPTTPGVALPWEAIAASRDPOFGIGIMETLIGADVVKLSGARAGALS 290
55 -----REIFSNLSIPAQRIGV-----ARHPRRSTLDYIQAISDWEIHLHGDRNGSD 101
291 PAVVALARIGRPRVVLIGOD-----RRFTL-GPQELRPARRGISLARLNLPIV 339
102 LALVGVALDGPVFLGHQGRDTONVLRNFGMASFGYRKALRLMEHADRFGMPL 161
340 SIIDTSGAELSGAAEELGIASSIARTLSKLIDAPLPTVSVIIIGGVGGALAMLPADLVY 399
162 TFDITPGAVAGVSAEELGGGGAIVANLRMEFRFSVPLICTVIGGGSGGALGIVGDRLL 221
400 AAENAMLSALPPEGASAILFRDTNHAETIERQGVQAHALLSGGLIDGIVA----- 450

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:40:30 ; Search time 366 Seconds
(without alignments)
244.909 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448
Sequence: 1 VKRPPYVWGMGHEHTSALT.....NNPERAGRDSRFRERLQA 491

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications, AA.*
2: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2448	100.0	491	US-10-024-370-3	Sequence 3, Appl1
2	2445	99.9	491	US-09-738-626-4429	Sequence 4429, Ap
3	2447	42.8	495	US-09-712-363-181	Sequence 181, App
4	487.5	19.9	465	US-10-156-761-13252	Sequence 13252, A
5	400	16.3	288	US-09-815-242-10925	Sequence 10925, A
6	384	15.7	289	US-09-815-242-11385	Sequence 11385, A
7	384	15.7	289	US-09-815-242-11385	Sequence 11385, A
8	375	15.3	330	US-09-815-242-13952	Sequence 13952, Ap
9	374	15.3	330	US-09-815-242-13952	Sequence 13952, A
10	368	15.0	296	US-09-815-242-11212	Sequence 11212, A
11	343.5	13.9	285	US-09-815-242-12618	Sequence 12618, A
12	340.5	13.9	285	US-09-815-242-12618	Sequence 12618, A
13	334	13.6	288	US-09-815-242-13618	Sequence 13618, A
14	334	13.6	288	US-09-815-242-13618	Sequence 13618, A
15	329.5	13.5	290	US-09-815-242-11898	Sequence 11898, A

16	291	11.9	532	15	US-10-156-761-10868	Sequence 10868, A
17	288	11.8	315	9	US-09-815-242-11014	Sequence 11014, A
18	286	11.7	531	15	US-10-156-761-10240	Sequence 10240, A
19	285	11.6	527	12	US-10-045-612A-31	Sequence 31, Appl1
20	277	11.3	309	9	US-09-815-242-5892	Sequence 5892, Ap
21	275	11.2	314	9	US-09-815-242-12908	Sequence 12908, A
22	275	11.2	526	15	US-10-187-267A-47	Sequence 47, Appl1
23	270	11.0	266	9	US-09-815-242-4964	Sequence 4964, Ap
24	269	11.0	319	9	US-09-815-242-13897	Sequence 13897, A
25	268.5	11.0	255	9	US-09-815-242-13366	Sequence 13366, A
26	268.5	11.0	255	9	US-09-815-242-13619	Sequence 13619, A
27	268.5	11.0	514	15	US-10-156-761-9840	Sequence 9840, Ap
28	266	10.9	319	9	US-09-815-242-10048	Sequence 10048, A
29	265.5	10.8	324	9	US-09-841-132-496	Sequence 496, App
30	265	10.8	262	9	US-09-815-242-10926	Sequence 10926, A
31	261.5	10.7	484	9	US-09-925-301-1140	Sequence 1140, Ap
32	261.5	10.7	484	15	US-10-106-698-5893	Sequence 5893, Ap
33	257	10.5	312	9	US-09-815-242-11339	Sequence 11339, A
34	257	10.5	312	9	US-09-815-242-11511	Sequence 11511, A
35	254	10.4	516	10	US-09-738-626-6665	Sequence 6665, Ap
36	252.5	10.3	324	9	US-09-841-132-566	Sequence 566, App
37	252.5	10.3	327	9	US-09-767-479-12	Sequence 12, Appl1
38	249.5	10.2	316	9	US-09-815-242-11934	Sequence 11934, A
39	241	9.8	537	10	US-09-738-626-4284	Sequence 4284, A
40	218.5	8.9	532	15	US-10-156-761-11401	Sequence 11401, A
41	217.5	8.9	593	15	US-10-224-539A-12	Sequence 12, Appl1
42	215.5	8.8	543	10	US-09-738-626-4285	Sequence 4285, Ap
43	207.5	8.5	543	15	US-10-259-648-2	Sequence 2, Appl1
44	190.5	7.8	563	15	US-10-224-539A-10	Sequence 10, Appl1
45	190.5	7.8	563	15	US-10-205-823-254	Sequence 254, App

ALIGNMENTS

```
RESULT 1
US-10-024-370-3
; Sequence 3, Application US/10024370
; Publication No. US20020142405A1
; GENERAL INFORMATION:
; APPLICANT: TILG, YOONNE
; APPLICANT: ELKMANN, BERND
; APPLICANT: EGGELING, LOTMAR
; APPLICANT: SAHM, HERMANN
; APPLICANT: MCKEL, BETTINA
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
; TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
; FILE REFERENCE: 21123-284139-MAS
; CURRENT APPLICATION NUMBER: US/10/024,370
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/362,899
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: DE 199 24 365.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO: 3
; TYPE: PRN
; ORGANISM: Corynebacterium glutamicum
US-10-024-370-3
Query Match 100.0%; Score 2448; DB 14; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.1e-217;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VKRPPYVWGMGHEHTSALTLDVDPDSFSNMTPEYDUNQGYATIERARKACD 60
DB 1 VKRPPYVWGMGHEHTSALTLDVDPDSFSNMTPEYDUNQGYATIERARKACD 60
QY 61 ESVTGEETGVIPIAVILSDPSFAGSLGVASVRIKAIHRAETELPLIVSPASGCA 120
```

Db 61 ESVITGEGTVEGIPVAVILSDSFSGSLGTVAVRIMKAIHRATELKLPLVSPASGA 120
QY 121 RMOEDNRAFWVMVSTTAQVQREAHLPVLYLNPMTGAMASWSSGHLTFAPGAOI 180
Db 121 RMOEDNRAFWVMVSTTAQVQREAHLPVLYLNPMTGAMASWSSGHLTFAPGAOI 180
QY 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVAKTLKVIQVEATDR 240
Db 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVAKTLKVIQVEATDR 240
QY 241 FSPPTPGVALPVMEAIARSRDPOPGIGEIMETLGADVVKLSGARAGALSPAVVALARI 300
Db 241 FSPPTPGVALPVMEAIARSRDPOPGIGEIMETLGADVVKLSGARAGALSPAVVALARI 300
QY 301 GGRPVVLIGQDPRFTLGPQELRFPARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360
Db 301 GGRPVVLIGQDPRFTLGPQELRFPARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360
QY 361 STARTLSKLIDAPLPTVSVIIIGQVGGGALAMLPAADVAAENAMLSALPEGASALIFR 420
Db 361 STARTLSKLIDAPLPTVSVIIIGQVGGGALAMLPAADVAAENAMLSALPEGASALIFR 420
QY 421 DTNHAABIIRQGVQAHALLSQGLIDGIVAFETHEFVEEIIIGTISNALSELNNPERAGRD 480
Db 421 DTNHAABIIRQGVQAHALLSQGLIDGIVAFETHEFVEEIIIGTISNALSELNNPERAGRD 480
QY 481 SRTFRFERLAQ 491
Db 481 SRTFRFERLAQ 491

RESULT 2

US-09-738-626-4429
; Sequence 4429, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4429
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4429

Query Match 99.9%; Score 2445; DB 10; Length 491;
Best Local Similarity 99.8%; Pred. No. 2e-217;
Matches 490; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEKRPYVWGMGMEHNSALTLLIDSVDDPSFISWNETPOYDNNNGVAYETLEBARAKACD 60
Db 1 MEKRPYVWGMGMEHNSALTLLIDSVDDPSFISWNETPOYDNNNGVAYETLEBARAKACD 60
QY 61 ESVITGEGTVEGIPVAVILSDSFSGSLGTVAVRIMKAIHRATELKLPLVSPASGA 120

Db 61 ESVITGEGTVEGIPVAVILSDSFSGSLGTVAVRIMKAIHRATELKLPLVSPASGA 120
QY 121 RMOEDNRAFWVMVSTTAQVQREAHLPVLYLNPMTGAMASWSSGHLTFAPGAOI 180
Db 121 RMOEDNRAFWVMVSTTAQVQREAHLPVLYLNPMTGAMASWSSGHLTFAPGAOI 180
QY 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVAKTLKVIQVEATDR 240
Db 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVAKTLKVIQVEATDR 240
QY 241 FSPPTPGVALPVMEAIARSRDPOPGIGEIMETLGADVVKLSGARAGALSPAVVALARI 300
Db 241 FSPPTPGVALPVMEAIARSRDPOPGIGEIMETLGADVVKLSGARAGALSPAVVALARI 300
QY 301 GGRPVVLIGQDPRFTLGPQELRFPARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360
Db 301 GGRPVVLIGQDPRFTLGPQELRFPARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360
QY 361 STARTLSKLIDAPLPTVSVIIIGQVGGGALAMLPAADVAAENAMLSALPEGASALIFR 420
Db 361 STARTLSKLIDAPLPTVSVIIIGQVGGGALAMLPAADVAAENAMLSALPEGASALIFR 420
QY 421 DTNHAABIIRQGVQAHALLSQGLIDGIVAFETHEFVEEIIIGTISNALSELNNPERAGRD 480
Db 421 DTNHAABIIRQGVQAHALLSQGLIDGIVAFETHEFVEEIIIGTISNALSELNNPERAGRD 480
QY 481 SRTFRFERLAQ 491
Db 481 SRTFRFERLAQ 491

RESULT 3

US-09-712-363-181
; Sequence 181, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-181

Query Match 42.8%; Score 1047; DB 10; Length 495;
Best Local Similarity 47.8%; Pred. No. 4.5e-88;

```

Matches      233; Conservative    73; Mismatches   153; Indels     28; Gaps
Oy          23 STLDPPDSFISNMETPOYDNLNOCYAETLERASAKCDESVITGEGTBCIPAVYLISDF 82
Db          12 AVLDKGSFVSWSEPLAVPADSYARELAAARATGADSDSVCOTGEGRVIGRRVAACFE 71
Oy          83 SFLGSGLGVAIVRIMKAHRAATELKLPLLVSASGAGMOENRPAFMMVSTIAVORH 142
Db          72 DFLGGSIGVAAERLTIAAVERATAERLPILASPSGGTMOEQTVAFLQWVKIAAIOLH 131
Oy          143 REAHLPFLVYLRNPRTMGAMASWGSSGHLLTFAPGAQIGFLGPRVVELTTGHALPDGYQC 202
Db          132 NQARLPYLVLRLHPPTTGCVAFASWMSGHLTVAEFGALIFELGRVVELIXGPDPFSGVOT 191
Oy          203 AENVLYKTGVIDGISVBLQIRAAVAKTLKYI---QPEVATDRSPPTPGVA--LPVMEAI 256
Db          192 AENVLRHGIIIDVVALLDRLRPMLDRALTJVLIALBEPLPA-----PQTPAVPVDVPMWSD 246
Oy          257 ARSRDPORPGIGEIMETTGADVUKLSGARAGALSAPRVVALAIRIGRPVVLVGODRRF-- 314
Db          247 VASRRPDPGVAYQLRHGATDRVLLSGTDGGBATTL-IALARFGSQPPIVVLGGQDAVGC 305
Oy          315 ---TLGPOELRFARRGISLARELNPIYSIIDTSGAELSQAABELGIASSIRTSLKIT 371
Db          306 GGSTYGPALRRARRRGMALEALCEPLVIVIDAAGALSNAAEQGLAQIHCLAEVLT 365
Oy          372 APPLPVSVYITIGCGVGAGALAMLPA DLVVAENANMLSALPREGASAILFRDTNNAEIER 431
Db          366 LDTPIVSILLGCGSGGPPALAMPADRVLNALHGMWLALPREGASAI VFDDTAHAABELAA 425
Oy          432 QGVQHALLSQGLDGIVAETHFVEILG---TISNALSELDN-----NEBRGRDR 482
Db          426 QGIRADLLKSQIVDTIVEPYDADEPLEFALRLSNALAEVHALRKIPAPERLA--TR 483
Oy          483 FTRPERL 489
Db          484 LQRYRRI 490

RESULT 4
US-10-156-761-13252
; Sequence 13252, Application US/10156761
; Publication No. US20030119018A1
GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13252
LENGTH: 465
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13252

Query Match      19.9%; Score 487.5; DB 15; Length 465;
Beat Local Similarity 29.9%; Pred. No. 2,4e-36;
Matches 136; Conservative 66; Mismatches 208; Indels 45; Gaps 7;

Oy          18 LTLLIDSVDDPPSFISNMETPOYDNLNOCYAETLERASAKCDESVITGEGTBCIPAV 77
Db          21 LPVAEGAAVAPGPCPLSM-----QGVDASRAARATERGEDSDSVCGRARIDGTPAVL 70
```

QY		78	ILSDPSFGSGSIGTAVASVIMKAIHRAETELKPLPVSPPSGARMOEDNRAFYMMWSITR	137
Db		71	IATFEGCFGSSIGERTGRLENAAYTAARAHLRPVSVLVTGSSRNQESGMALTQLORVAR	130
QY		138	AVQRREAHLPVLVYLNRNPWTMGAMASWSSGHLTFAEBGAOIGFLPPRVVELTTGHALP	197
Db		131	QSALTRRAGLPQALVRDPTTGCGMATLGAGADVLLALPGVOVGAFGSRV-----RRP	183
QY		198	D---GVQAENLVKTGVLDIGIVSPLOLRPAVAKITKVIOPEVATRDFSTTFRGVALPWME	254
Db		164	DADPAAYTRAQOLAAGSIDAVVRPOELRETVEGLMALL-----TAPSTEPAAPPALGTSD	239
QY		255	-----AIRSRDPORGIGEIMETIGADVVKLSGARAGALSAPVRVAL-----	297
Db		240	RLPDTGMAVAGRAREPERPRAOAYIDAYFTTRAAIAGRCGGTDGMLCGFGFSGSSGSG	299
QY		298	--ARIGRPVLLIGODRRFFTLGPOLBFPARRGISLARELNPIVISIIDTSGBNELSOAME	354
Db		300	SGSGEHGGRVVAYAQTGTAT--RPAGYRTATRLILADBLGIPLTVLTVPGANDAEAE	358
QY		355	ELGIASSIARTLSKIDNP.LPVSIIITOGVGGGALMLPALDVYAANAEMASALPREGA	414
Db		359	ROGVGAATAIEIFGAVATRTPLPTILLIBGSGGALALAAPGNWATPDSPSYVIAPELA	418
QY		415	SAILFRDTHAAEIERQGVONHALLSOGLDIGIV	449
Db		419	AAILKRPEEARPTRADQLRVAPQDLVALGVVRGIV	453
 RESULT 5 US-09-815-242-10925 Sequence 10925, Application US/09815242 Patent No. US20020061569A1 GENERAL INFORMATION: APPLICANT: Haselbeck, Robert APPLICANT: Ohlsen, Karl L. APPLICANT: Zykkind, Judith W. APPLICANT: Wall, Daniel APPLICANT: Trawick, John D. APPLICANT: Carr, Grant J. APPLICANT: Yamamoto, Robert T. TITLE OF INVENTION: Identification of Essential Genes in FILE REFERENCE: ELITRA.011A CURRENT APPLICATION NUMBER: US/09/815,242 CURRENT FILING DATE: 2001-03-21 PRIOR APPLICATION NUMBER: 60/191,078 PRIOR FILING DATE: 2000-03-21 PRIOR APPLICATION NUMBER: 60/206,848 PRIOR FILING DATE: 2000-05-23 PRIOR APPLICATION NUMBER: 60/207,727 PRIOR FILING DATE: 2000-05-26 PRIOR APPLICATION NUMBER: 60/242,578 PRIOR FILING DATE: 2000-10-23 PRIOR APPLICATION NUMBER: 60/253,625 PRIOR FILING DATE: 2000-11-27 PRIOR APPLICATION NUMBER: 60/257,931 PRIOR FILING DATE: 2000-12-22 PRIOR APPLICATION NUMBER: 60/269,308 PRIOR FILING DATE: 2001-02-16 NUMBER OF SEQ ID NOS: 14110 SOFTWARE: PasteSeq for Windows Version 4.0 SEQ ID NO 10925 LENGTH: 288 TYPE: PRT ORGANISM: Enterococcus faecalis US-09-815-242-10925				
Query Match	16.3%	Score 400;	DB 9;	Length 288;
Best Local Similarity	42.6%;	Pred. No. 1,4e-28;		
Matches	89;	Conservative	39;	Mismatches 79;
				Indels 2;
				Gaps 2;

```
OY 25 LDPDSFISW-NETPOYDNLN-OGVAETLERARSKAKCDBESVITGEGTVEGIPVAVILSDPF 82
DB 73 VDEKSFHNMDSLTKDPLNPFPGYIEKIKMOEKGLEBAVLTGATTEGCAVVAIGCIDA 132
OY 83 SFLGSLGTVA SVIRIMKAIHRATELKLPLVSPASGARMODNRAFMVMSITAAVORH 142
DB 133 NPMGSMGTIVGEEKITRLFERATEKHLFPVIFTASGARMOGIFSLMQMAKISAAQRH 192
OY 143 REAHLPFLVLRNPTMGAMASWGSSGHLTFAPGAQIGFLGPRVVELTTGHALPDGVOQ 202
DB 193 NKGGLLYLVLTPTGTGGVATSPADMDGIIILPEPSLIGFAGRRIYEQIRROELPDDPOK 252
OY 203 AENLVKTVIGDGVSPLOLRAAVAKTLKV 231
DB 253 AEFLEHGFVDQIVPRNLIKORLSLRL 281

RESULT 6
US-09-815-242-11385
; Sequence 11385, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11385
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11385

Query Match 15.7%; Score 384; DB 9; Length 289;
Best Local Similarity 39.3%; Pred. No. 4.3e-27;
Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;
```

```
OY 206 LVKTGVIDGVSPLOLRAAVA-----KTLKV 231
DB 255 LLEHGLIDMTVHRKDLKKTLSDLIAMTHTKTSKI 288

RESULT 7
US-09-895-913A-98
; Sequence 98, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE REFERENCE: 06132/043002
; CURRENT FILING DATE: 2001-06-29
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-98

Query Match 15.7%; Score 384; DB 10; Length 289;
Best Local Similarity 39.3%; Pred. No. 4.3e-27;
Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;
```

```
OY 26 DPDSFISWNETPOYDNLNOCVAETLERARSKAKCDBESVITGEGTVEGIPVAVILSDPSFL 85
DB 85 DPLNFVD-----KESYKQRIKKYEKRTNRPSVIGSEAKINMPLOIVFDFSF 134
OY 86 GSGISGTVA SVIRIMKAIHRATELKLPLVSPASGARMODNRAFMVMSITAAVORHREA 145
DB 135 GSGISVGEKIVRAINRAVAKREALIVASGARMODESTYSLMQMAKISAAQLRLSEA 194
OY 146 HLPFLVLRNPTMGAMASWGSSGHLTFAPGAQIGFLGPRVVELTTGHALPDGVOQAE 205
DB 195 KLPPISLSLSDPTGGVATSPADMDGIIILPEPSLIGFAGRRIYEQIRROELPDDPOK 254

OY 206 LVKTGVIDGVSPLOLRAAVA-----KTLKV 231
DB 255 LLEHGLIDMTVHRKDLKKTLSDLIAMTHTKTSKI 288

RESULT 8
US-09-815-242-13952
; Sequence 13952, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13952
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13952
```

```

Query Match      15.3%; Score 375; DB 9; Length 330;
Best Local Similarity 36.0%; Pred. No. 3,6e-26;
Matches 85; Conservative 49; Mismatches 92; Indels 10; Gaps 3;
```

```

Qy 17 ATTLDVLDPDSFISWNETPOY-DNLNGVAETLERARSKAKCDESVITGEGVGPV 75
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 SLVEIGSELEPDKVLFKRSKYYKORLASAQKETEK-----DALVVMKGTLHGMPV 146
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 76 AVILSDFSFLGSLGTVA SVIRIMKAIHRATELKLPLVSPASGARMQEDNRAFYVMVSI 135
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 VAAAEFPAFMGSGMSVVGARFVRAVEQALDENCPICFSASGARGMQLMSLMQMAKT 206
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 136 TAAVGRHREAHLPFLVYLRNPTMGAMASWGSSGHLTPAEPQAOTGFLGPRVVELTTGHA 195
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 SAAAKMQERGLPYISVLTDPMTGVSASFAMLGDLNIAEPALIGFAGPRVIEQTVREK 266
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 196 LPDGVQAENLVKTVGVIDGIVSPLQLRAAVAKTLKVIQVEATDRFSPTTGVALP 251
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 LPPGRQSEPLEKKAIDMIVRRPRLKLSILAKLMLPAPNDAP-REGVVVP 321
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 9
US-09-815-242-10230
; Sequence 10230, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Twick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10230
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10230
```

```

Query Match      15.3%; Score 374; DB 9; Length 304;
Best Local Similarity 36.0%; Pred. No. 3,9e-26;
Matches 85; Conservative 49; Mismatches 92; Indels 10; Gaps 3;
```

```

Qy 17 ATTLDVLDPDSFISWNETPOY-DNLNGVAETLERARSKAKCDESVITGEGVGPV 75
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 SLVEIGSELEPDKVLFKRSKYYKORLASAQKETEK-----DALVVMKGTLHGMPV 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 76 AVILSDFSFLGSLGTVA SVIRIMKAIHRATELKLPLVSPASGARMQEDNRAFYVMVSI 135
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 VAAAEFPAFMGSGMSVVGARFVRAVEQALDENCPICFSASGARGMQLMSLMQMAKT 180
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 136 TAAVGRHREAHLPFLVYLRNPTMGAMASWGSSGHLTPAEPQAOTGFLGPRVVELTTGHA 195
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 SAAAKMQERGLPYISVLTDPMTGVSASFAMLGDLNIAEPALIGFAGPRVIEQTVREK 240
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 196 LPDGVQAENLVKTVGVIDGIVSPLQLRAAVAKTLKVIQVEATDRFSPTTGVALP 251
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 LPPGRQSEPLEKKAIDMIVRRPRLKLSILAKLMLPAPNDAP-REGVVVP 295
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 10
US-09-815-242-11212
; Sequence 11212, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Twick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11212
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11212
```

```

Query Match      15.0%; Score 368; DB 9; Length 296;
Best Local Similarity 36.8%; Pred. No. 1,3e-25;
Matches 81; Conservative 45; Mismatches 80; Indels 14; Gaps 3;
```

QY	25	DDPDSFISMNENPOY-DNUNGVAEETLEARSKACDEASTTEEGVEIGPVAIIISDFS	83
Db	79	LEPDKILFKDKKTKYDRINAAOKETEK-----DALITTTGTLINMPIVVAASNFA	130
QY	84	FLGSLGTAVASVRIKAIHRATELKLPLVSPASGARMQEDNRPAVMMVSIITPAVQRHR	143
Db	131	FMGSGMSGVGAKFYKAAEKAMENNCPPYCFOSFASGARQEBALFSLMQAKTSAYVLAQNR	190
QY	144	EAHLFPVLYLRNPTMGAMASWSSGHLTFPAEPQAOIGFLGPRVVELTTGHALPDGVOQA	203
Db	191	EKGVPFISVLTDPPTTGVSASFAMLDENLIAEKAALIGFPGPVIQVIREKLPBGFORS	250
QY	204	ENLVKTVYIDGVISPLQLEPAAVAKTLKVIQPVATEDRFSP	243
Db	251	EFLEKGAIDMLVKRSEMRQTIAASVLS-----KLITQPSB	285

```

RESULT 11
US-09-815-242-12618
; Sequence 12618, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12618
; LENGTH: 285
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12618

```

	Query Match	Similarity	14.0%;	Score 343.5;	DB 9;	Length 285;
	Best Local	82;	Conservative	47;	Mismatches	94;
					Indels	5;
					Gaps	3.
Qy	13	EHTSALT---	IDSVLDPDPSFISNMETPOYDN-LN-OGYAETLERARSKAKCDESVIGE	67		
		:::	:::	:::	:::	:::
Db	56	DHIALTYKRIEALSDBSGFTEFPKGMNSANPLDPFPTIEKLEKDDQKTKGAKENAVVGT	115			
Qy	68	GTVEGI PVAVILSDPSFLGSLGTVAASVRIMKAHRAELKLPILVSPASGARGMOEDNR	127			
		:::	:::	:::	:::	:::
		:::	:::	:::	:::	:::
Db	116	AQLDMEKGVAVMDSEFRMGSMGIVGEKICRIIDYCTENRPLPFLIFSAGARMOEGII	175			
Qy	128	AFVMMVSTTANQQRREALPLVLYLNRTMGGAAMSSGHLFAEKGQITGLGRV	187			
		:::	:::	:::	:::	:::
		:::	:::	:::	:::	:::
Db	176	SIMQKGTSTVSKRISDGLLYSVLTHTTGVGASAFSVDINISSEKALIGPARRV	235			

```
QY      188 VELTGHALBEGVQAENLVKTVGIDGISPLQLRAAVAKTLKVQPV 235
       :| | | | | | | | | | | | | | | | | | | | | | |
Db     236 IEQTINEKEPDDFQTAEPFLLEHGQLDKVHRNDMRQGLTSELILKIHQEV 283
```

```

RESULT 12
US-09-815-242-5424
: Sequence 5424, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 1410
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5424
: LENGTH: 252
: TYPE: prt
: ORGANISM: Staphylococcus aureus
US-09-815-242-5424

```

```

Query Match          13.9%: Score 340.5; DB 9; Length 255;
Best Local Similarity 35.8%: Pred. No. 3.7e-23;
Matches      81: Conservative      47; Mismatches    93; Indels      5; Gaps      3.

QY      13  EHTSALTLP---IDSLVDPDSFISWNMETPQYDN-IN-QGYAETLERARSKACDSEVTIGE 67
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      27  DHHIALTRKYKLEAISDBGSEFTTERDKGMTSANPLDFPEYIKIRKDOOKTGKAEAVVTGT 86
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      68  GTVEGIPYAVVILSDPSFLGSLGTVASVYRKAIHRALETLPLVSPASGARNQOEDNR 127
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      87  AQLDGMKRGVAVMDSRFRMSGVSGVIGEKIRIIDCYENNLPLFLFSASGARNQOECII 146
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      128  AFVMMVSTIAVQRHREAHPLVLYLRNPIMGAMASGSGGHLTPAEPGQIGFLGRV 187
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      147  SLMQGKTSVSLKRRSDAGLLIYLYLTPPTGGVASPASVDINLSEPKALIGPAGRRV 206
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      188  VELTTGHALPDGVQDAENLVKTVGYIDGVSPLOLRAAVAKTLKYIQ 233
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      207  IEQITNEKLPDDFQTAEFLLEHGOLDKVRHNDMRQTLSELIKIHQ 252
          ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-815-242-13365
; Sequence 13365, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.

```

```
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13365
LENGTH: 288
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13365
```

```
Query Match      13.6%; Score 334; DB 9; Length 288;
Best Local Similarity 39.0%; Pred. No. 1.8e-22;
Matches 73; Conservative 36; Mismatches 78; Indels 0; Gaps 0;
```

```
QY 45 GAAETLERARSKACDESVITGEGTVEGIPVAVILSDPSFLGSGIGTAVSRIMKAHRA 104
DB 96 GYOKKLAMREKRTGHEHENVITGALIKQTVAGLIMDSFTWASNGTVGKRTLFEYA 155
QY 105 TELKPLIVSPASGARMOEDNRAFMVMSITAAVORREAHLPVLVLRNPTMGAMAS 164
DB 156 TVEKLPVLFASGARMOEGIMSLMOMAKISAAYKHSNAGLPVLITLPTTGCVTAS 215
QY 165 WGSCHLTFAPPGAIOIGLGRVVELTGHALPDGVOAENLVKTVGIDGIVSPLOQRAA 224
DB 216 FAMEDDILAPQSLVGFAGRVIENTVRESLPEDFOKAEFLLEGHGFVDALVKRDLPT 275
QY 225 VAKTLKV 231
DB 276 IASLVRL 282
```

```
RESULT 14
US-09-815-242-13618
Sequence 13618, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
```

```
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13618
LENGTH: 288
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13618
```

```
Query Match      13.6%; Score 334; DB 9; Length 288;
Best Local Similarity 39.0%; Pred. No. 1.8e-22;
Matches 73; Conservative 36; Mismatches 78; Indels 0; Gaps 0;
```

```
QY 45 GAAETLERARSKACDESVITGEGTVEGIPVAVILSDPSFLGSGIGTAVSRIMKAHRA 104
DB 96 GYOKKLAMREKRTGHEHENVITGALIKQTVAGLIMDSFTWASNGTVGKRTLFEYA 155
QY 105 TELKPLIVSPASGARMOEDNRAFMVMSITAAVORREAHLPVLVLRNPTMGAMAS 164
DB 156 TVEKLPVLFASGARMOEGIMSLMOMAKISAAYKHSNAGLPVLITLPTTGCVTAS 215
QY 165 WGSCHLTFAPPGAIOIGLGRVVELTGHALPDGVOAENLVKTVGIDGIVSPLOQRAA 224
DB 216 FAMEDDILAPQSLVGFAGRVIENTVRESLPEDFOKAEFLLEGHGFVDALVKRDLPT 275
QY 225 VAKTLKV 231
DB 276 IASLVRL 282
```

```
RESULT 15
US-09-815-242-11898
Sequence 11898, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
```

```

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 11898
; LENGTH: 290
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
; US-03-815-242-11898

```

Query Match	13.5%	Score 329.5	DB 9	Length 290
Best Local Similarity	35.8%	Pred. No. 4.8e-22		
Matches 78	Conservative 39	Mismatches 88	Indels 13	Gaps 3

```

QY      25 .DDPDSFISNNENPOY -DNINQVAFELTEBARSKACXDESYTGEQVVEGIPAVVILSDFS 83
Db      81 LEVDURLFRDSKKYKIDRLAAAKOTDGEK-----DALIAMSGLQGFVACAFEFIS 132
QY      84 FLGSLGTVAASVRIMKAIHRATELKLPLIVSPASGARMQEDNRPAFVMMVSIYAAVQRHR 143
Db      133 FMGSGMGAIVGBRFPRAANVALEKRCPLCFSAISGARRQGEALISLMQAKTISAVLARLR 192
QY      144 EAHLEPFLVYLRNPPTMGAMASWSSGHLTFABEQAQIGLGPVVELTGHALPDGVOQA 203
Db      193 BEGIFPVSVLTDPVYGVSASLAMDGVIVGEKALIGFAGPVRIVEQTVAREKLPBGFORS 252
QY      204 ENLVYKTGVDIGVSPLOLRPAAVAKTLKVI-----QPVBA 237
Db      253 EFLLEHGALDMIVHRAELRPRLANLLISAFTHSPSPVA 290

```

Search completed: November 13, 2003, 06:57:37
Job time : 369 secs